AI-Stack Express

Machine Learning Management Platform

User Manual

Version：4.4

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

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

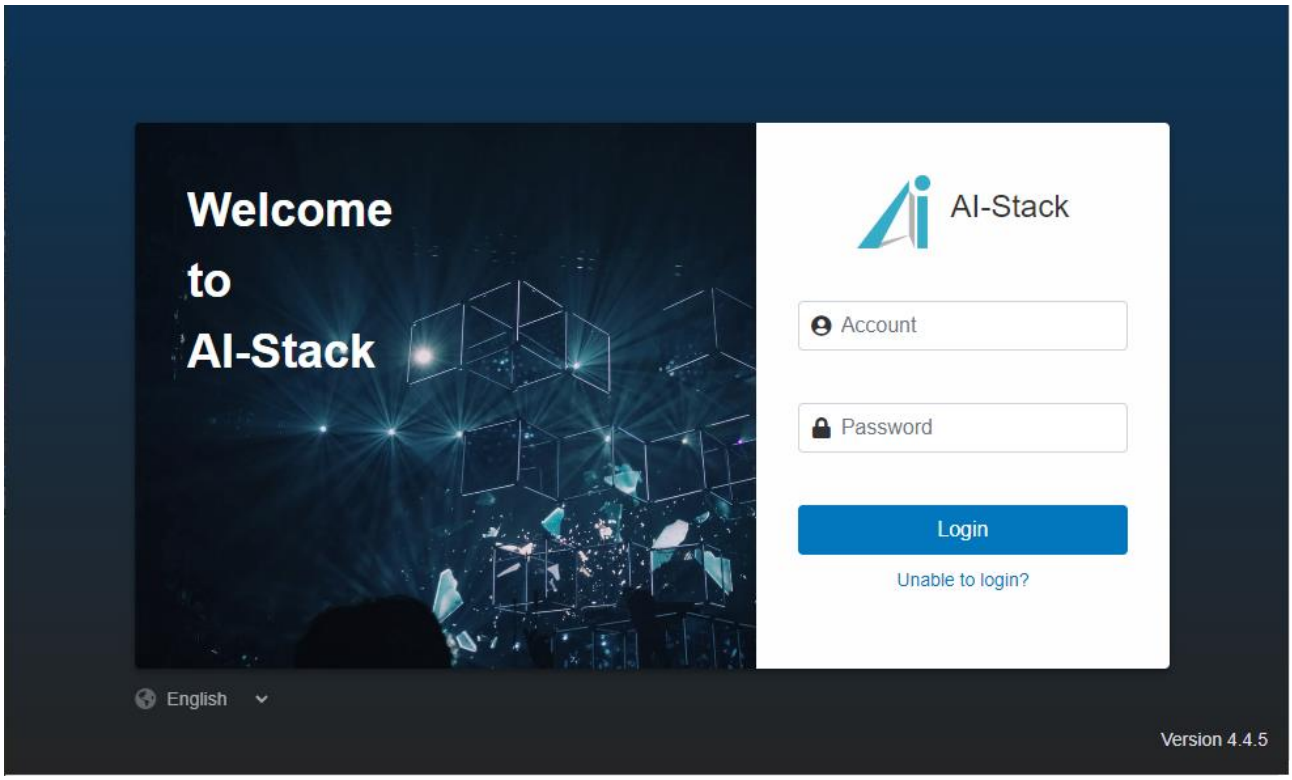
This manual introduces the features and the standard operation of AI-Stack – a Machine Learning Management Platform for platform users. It is recommended to use Chrome versions 60 or above / Firefox versions 60 or above. If the browser version fails to meet the requirements, the system may not function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

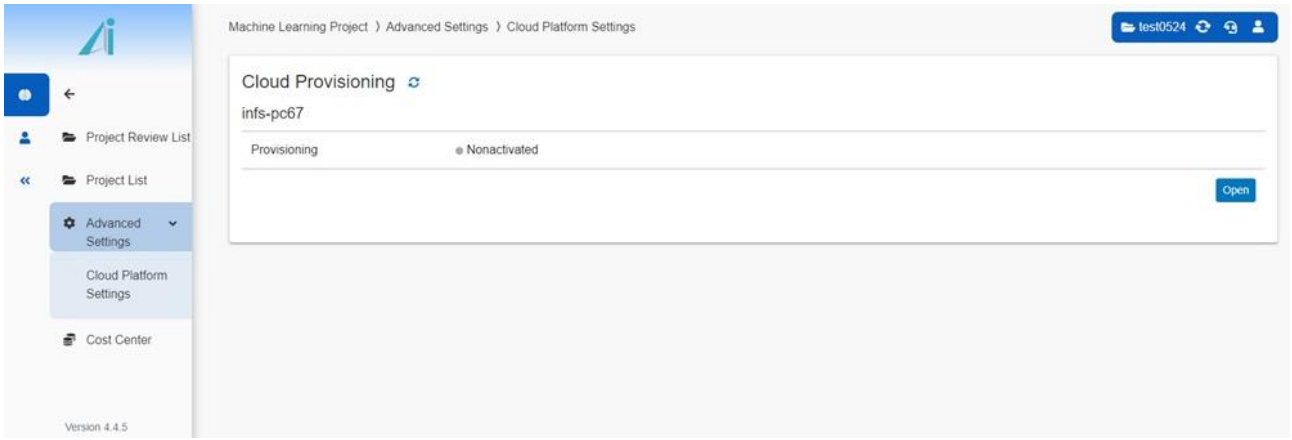
The system operates and organised based on projects which allows users to create project according to specific project goals, departments, small-scale experiments, internal competitions, educational training, event showcases, and various other usage scenarios. This differentiation of projects supports 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 plat form.

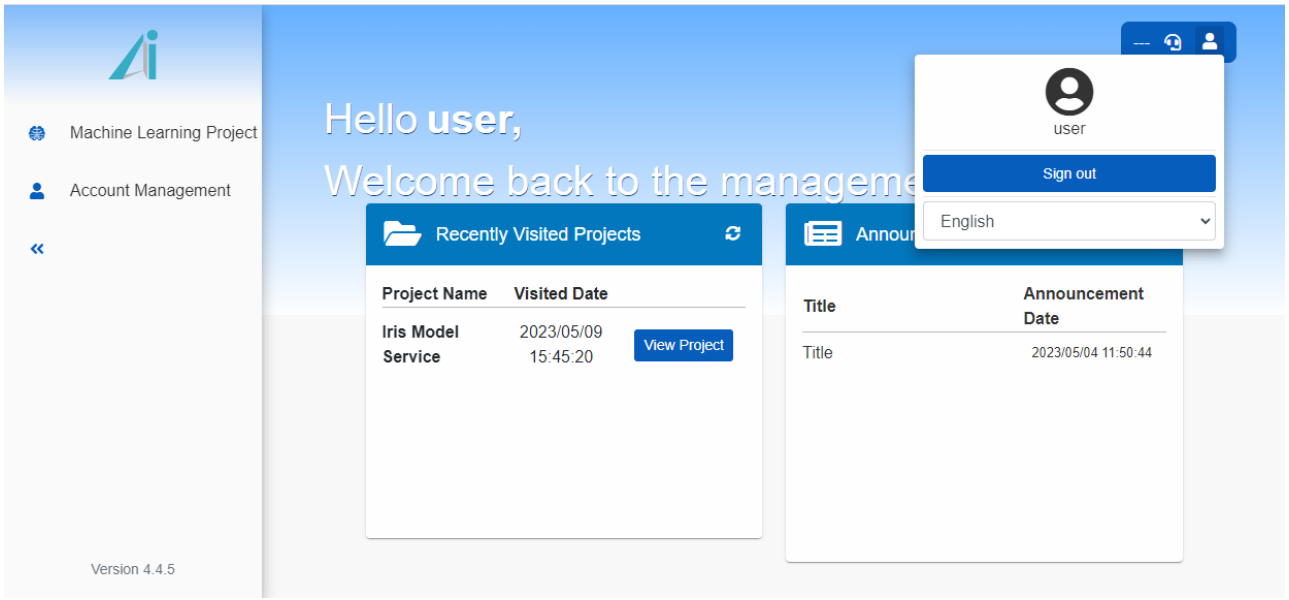


As the platform resources were not activated the first time you logged in, you will land at “Cloud Provisioning” page. To activate the resources, please refer to Section 4 Service Ac tivation.



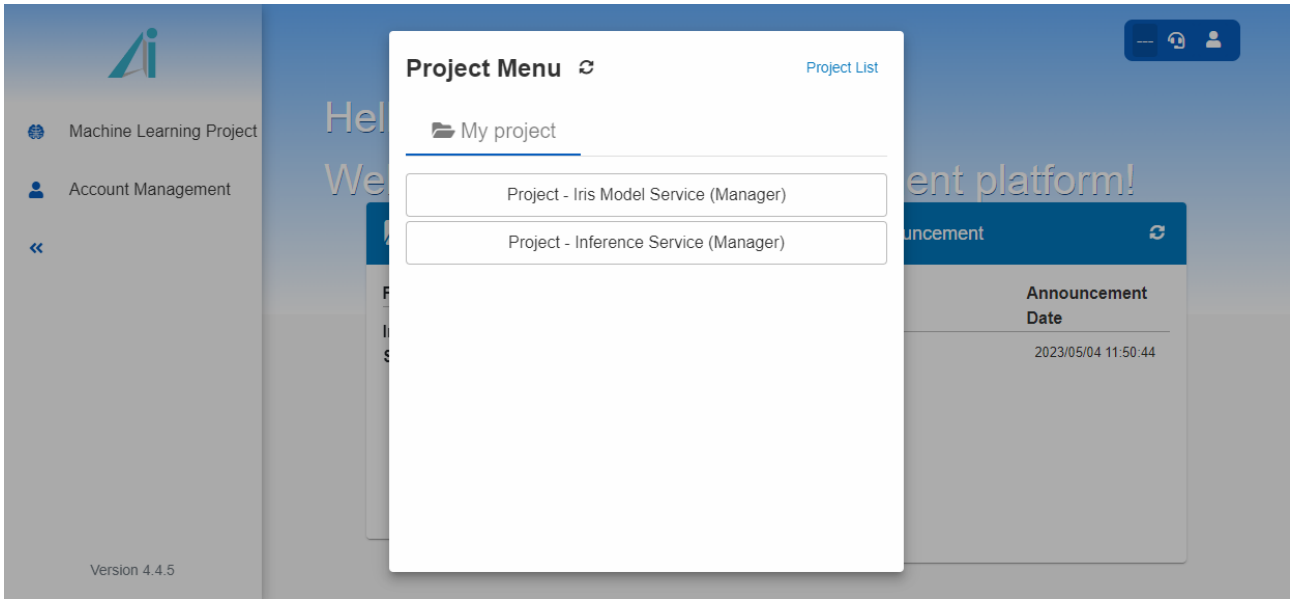
3. Interface

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

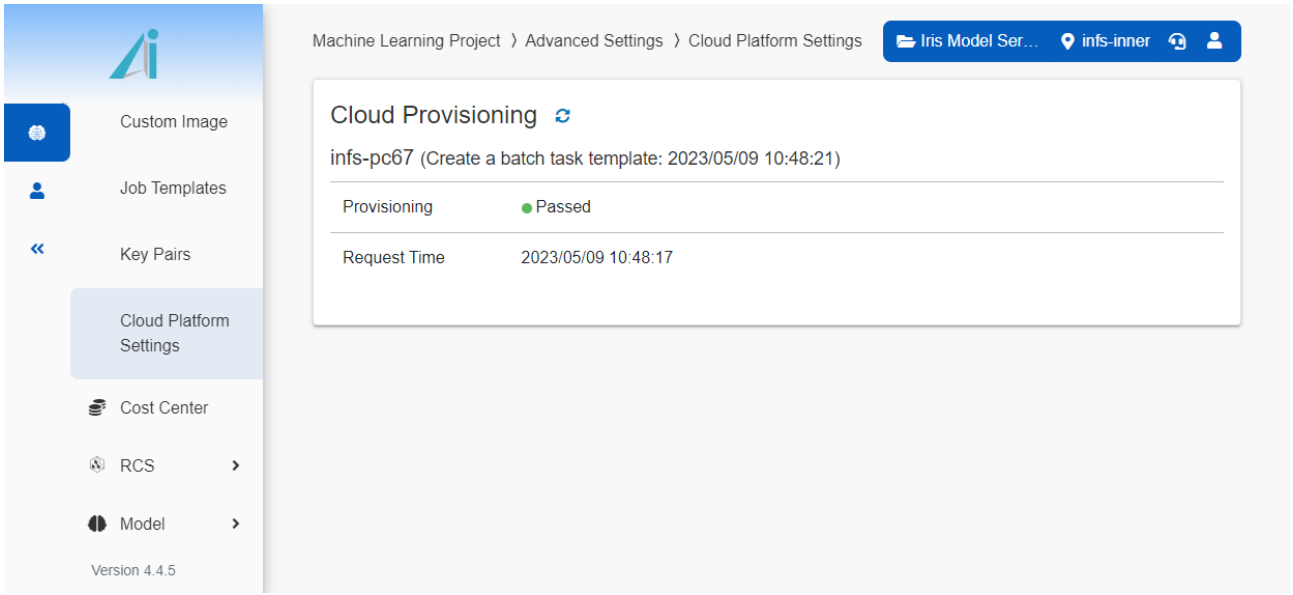
ojects with different purposes.



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 displaye d 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 t o complete.

Click to refresh the page until the status becomes “Passed”, and you can start usi ng our services.

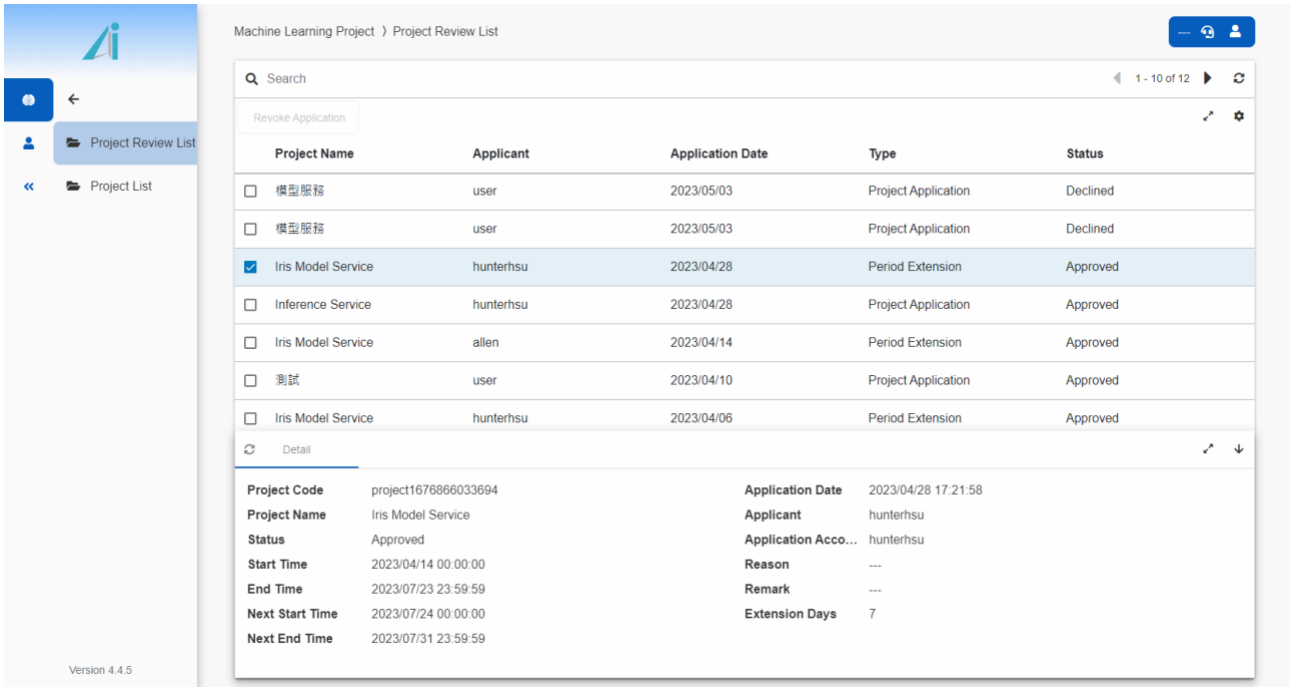

Custom Image
Job Templates
Key Pairs
Cloud Platform Settings
Cost Center
RCS
Model
Version 4.4.5
Machine Learning Project 〉Advanced Settings 〉Cloud Platform Settings
Iris Model Ser…
infs-inner
Cloud Provisioning
infs-pc67 (Create a batch task template: 2023/05/09 10:48:21)
Provisioning
Passed
Request Time
2023/05/09 10:48:17

5. Machine Learning Project

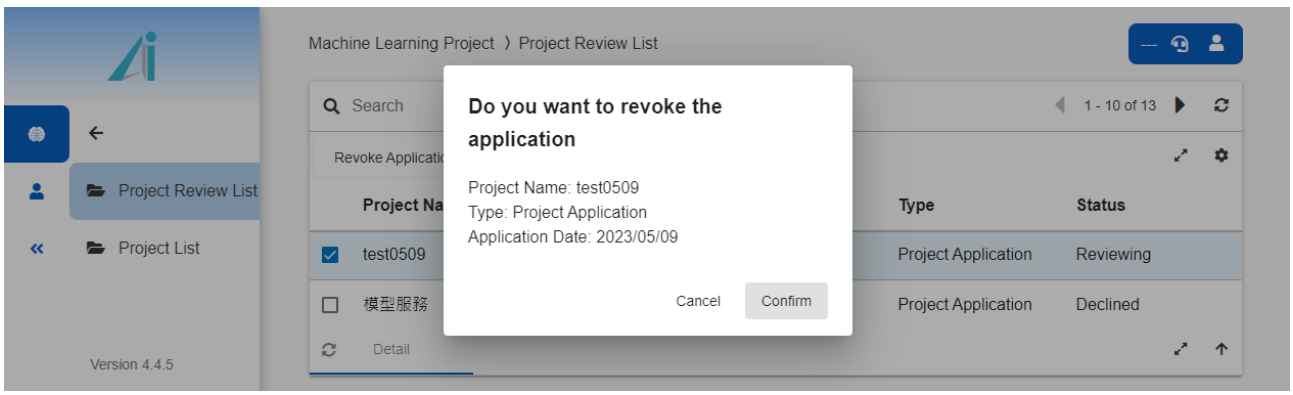
Users can use the project overview page to view project general information and resource usage status for each project. Projects have start and end dates, and prior to the project end date, the project manager will receive email notifications to extend the project. If the project is not extended after 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.



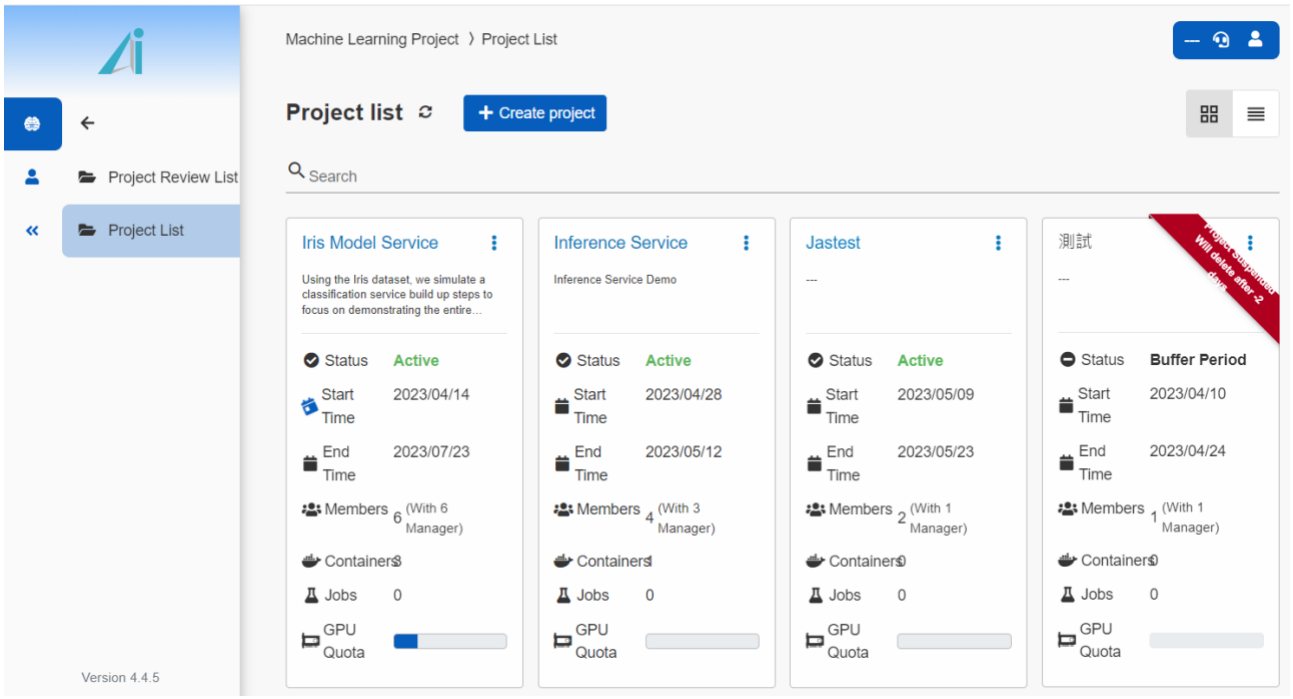
User can choose to revoke the application if the application status is still at “reviewing



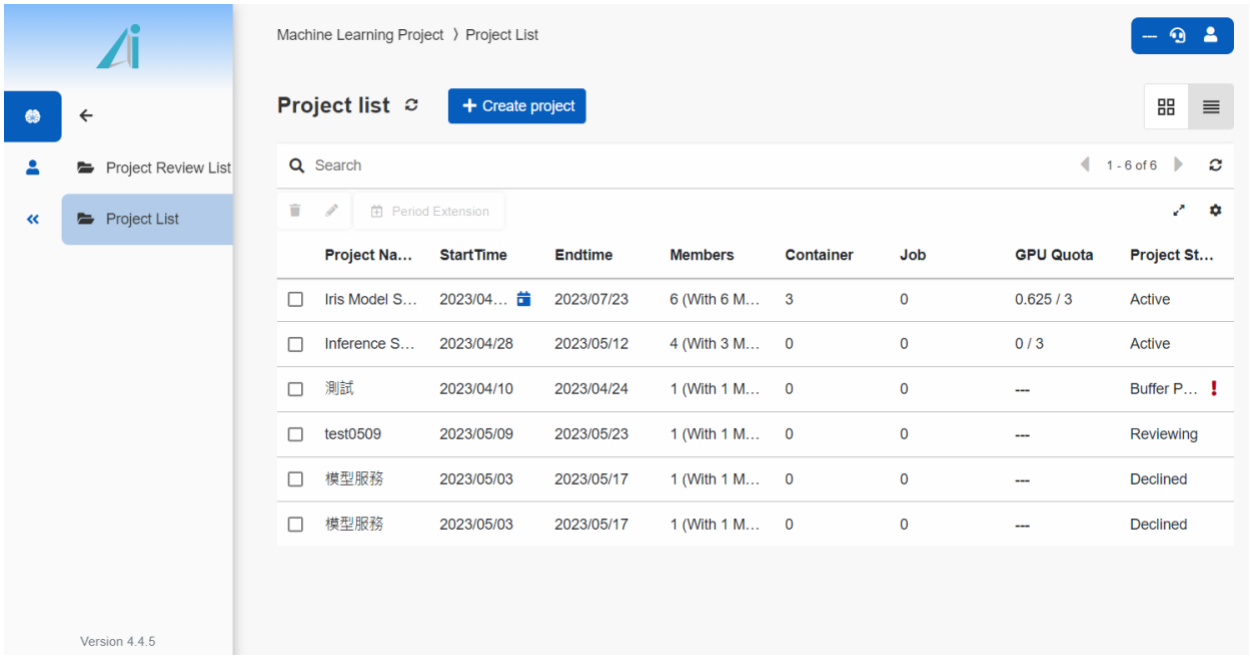
2. Project List

A. Project Overview

Clicked on project list to inspect the overall status of the projects. Default to card views, 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.


Machine Learning Project 〉Project List
Project list
+ Create project
Search
1 - 6 of 6
Period Extension
Project Review List
Project List
Version 4.4.5
Project Na...
StartTime
Endtime
Members
Container
Job
GPU Quota
Project St...
Iris Model S...
2023/04...
2023/07/23
6 (With 6 M...
3
0
0.625 / 3
Active
Inference S...
2023/04/28
2023/05/12
4 (With 3 M...
0
0
0 / 3
Active
測試
2023/04/10
2023/04/24
1 (With 1 M...
0
0

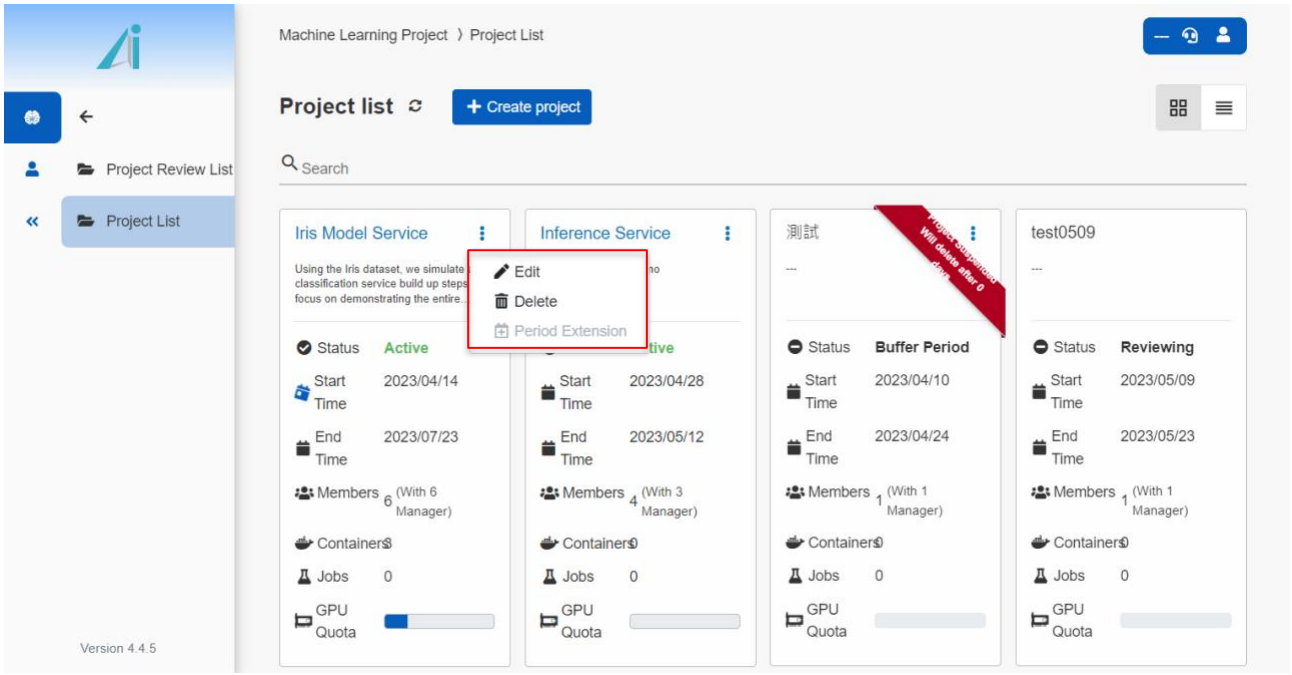
Buffer P...
test0509
2023/05/09
2023/05/23
1 (With 1 M...
0
0

Reviewing
模型服務
2023/05/03
2023/05/17
1 (With 1 M...
0
0

Declined
模型服務
2023/05/03
2023/05/17
1 (With 1 M...
0
0

Declined

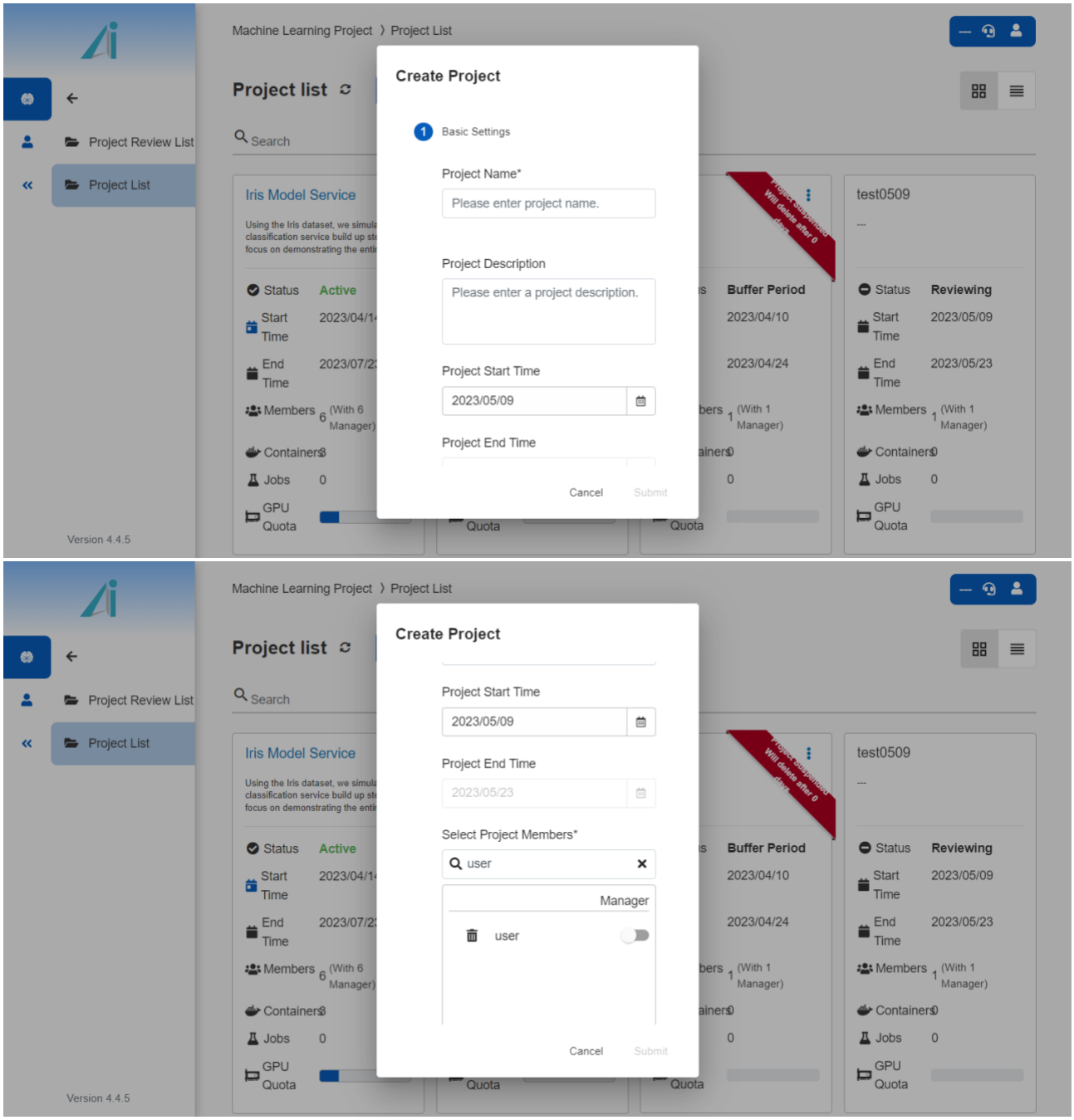
The project manager can edit a project, delete a project, or extend the project duration by clicking the ⋮ icon on the top right of each project card.



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)

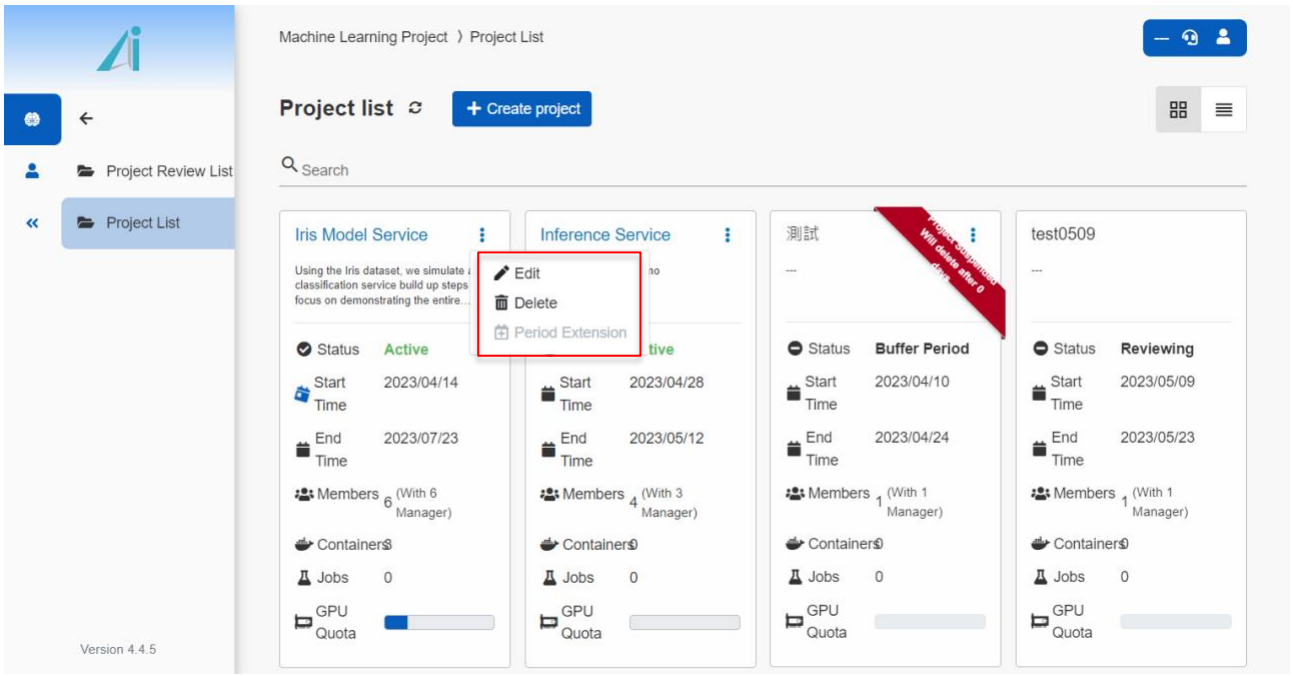


※ If the project duration was not extended after the buffering period, the platform wil l release the all the resources allocated to this project, which includes but not limited t o containers, job, and storage cluster.

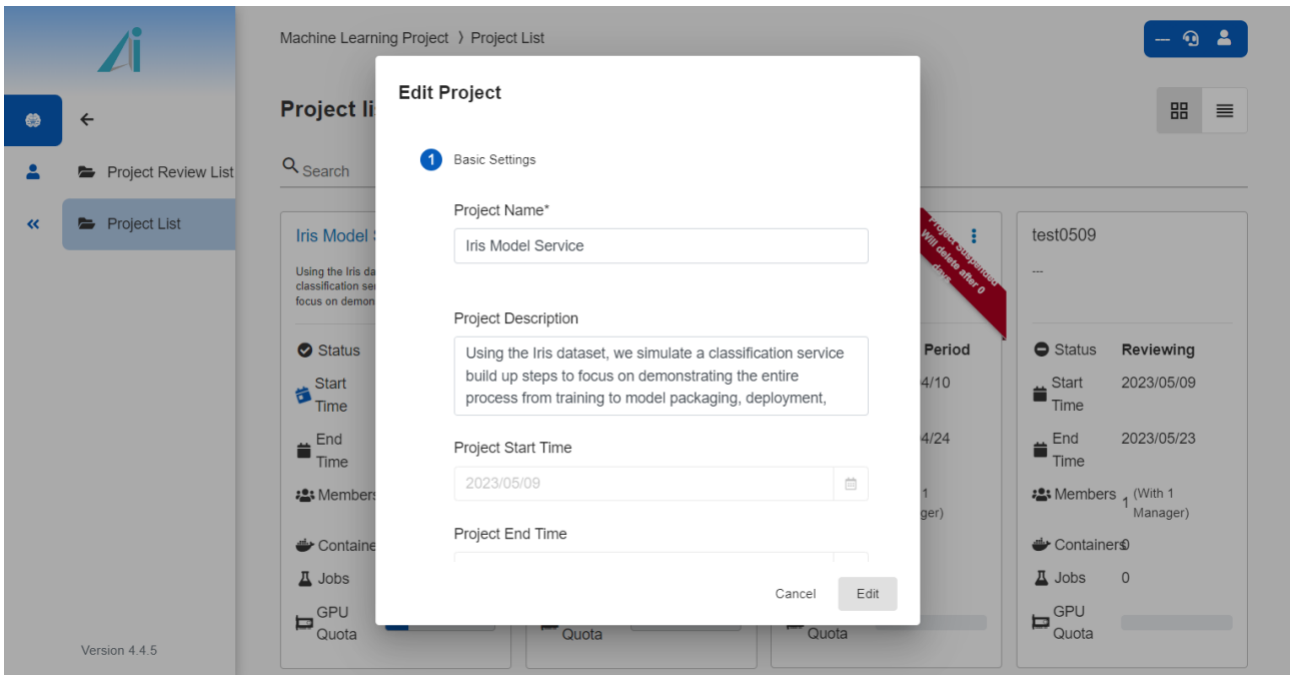
The applicant will receive an email notification such as the below figure once the proj ect application is approved.

C. Edit Project

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

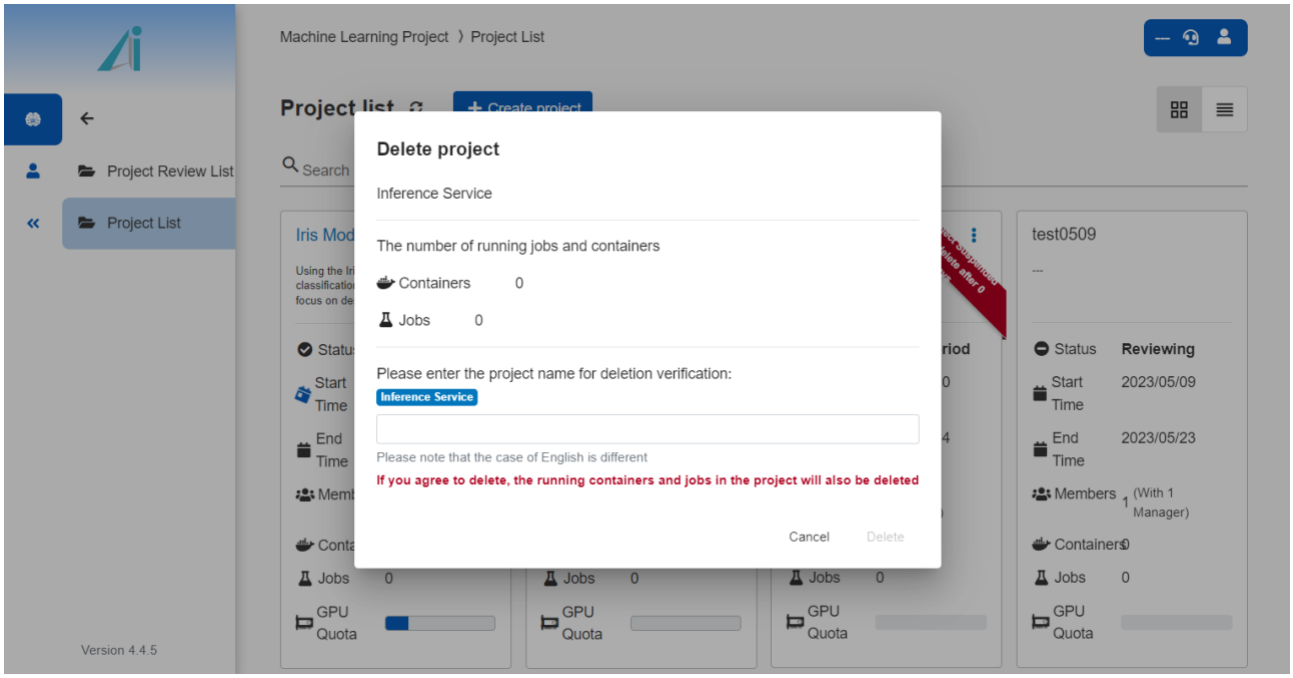


Project managers can edit project name, edit project description, add project members and assign other project managers.



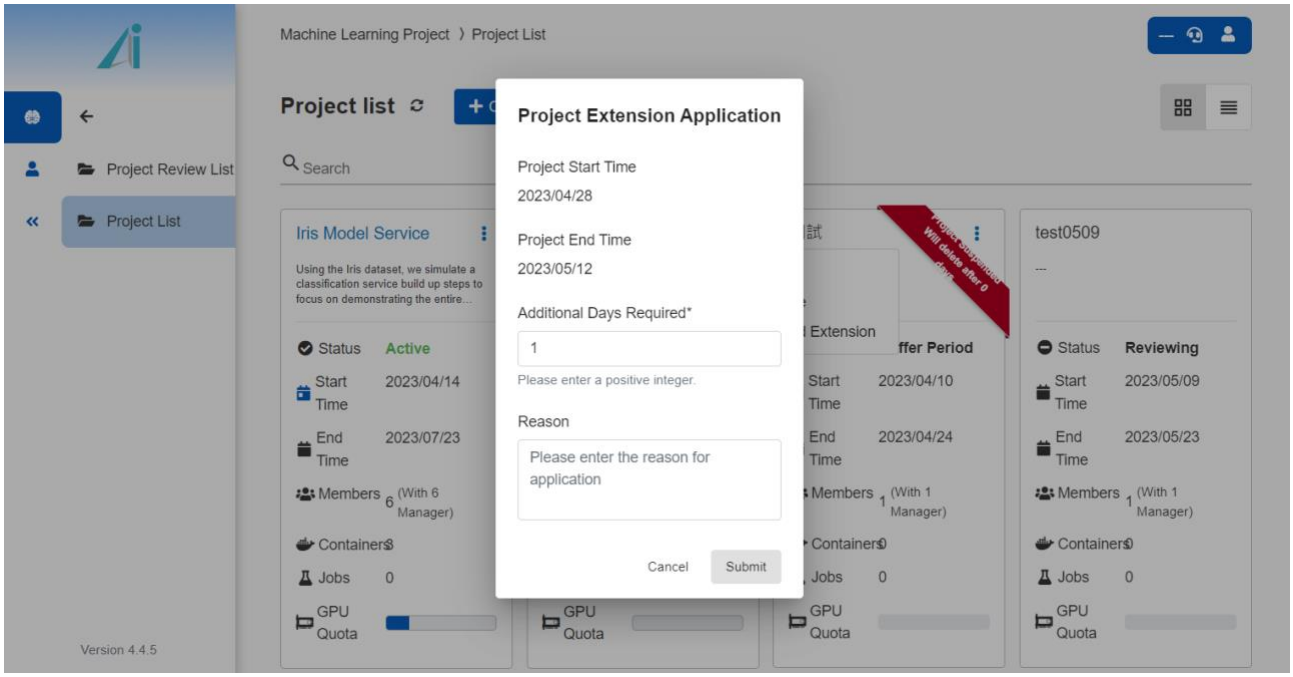
D. Delete Project

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

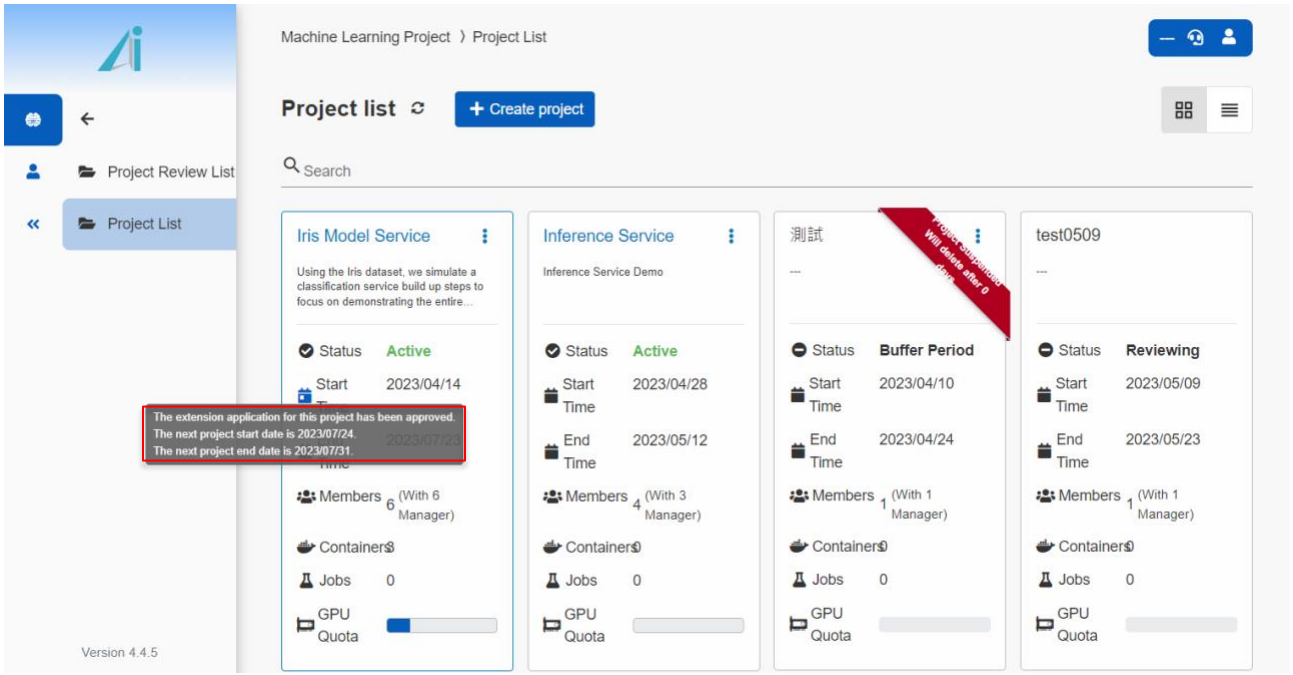


E. Project Extension Application

Project managers can extend the project duration by clicking on the ⋮ icon and pres s extension. Submit the application through the popup window after you enter the da ys required.

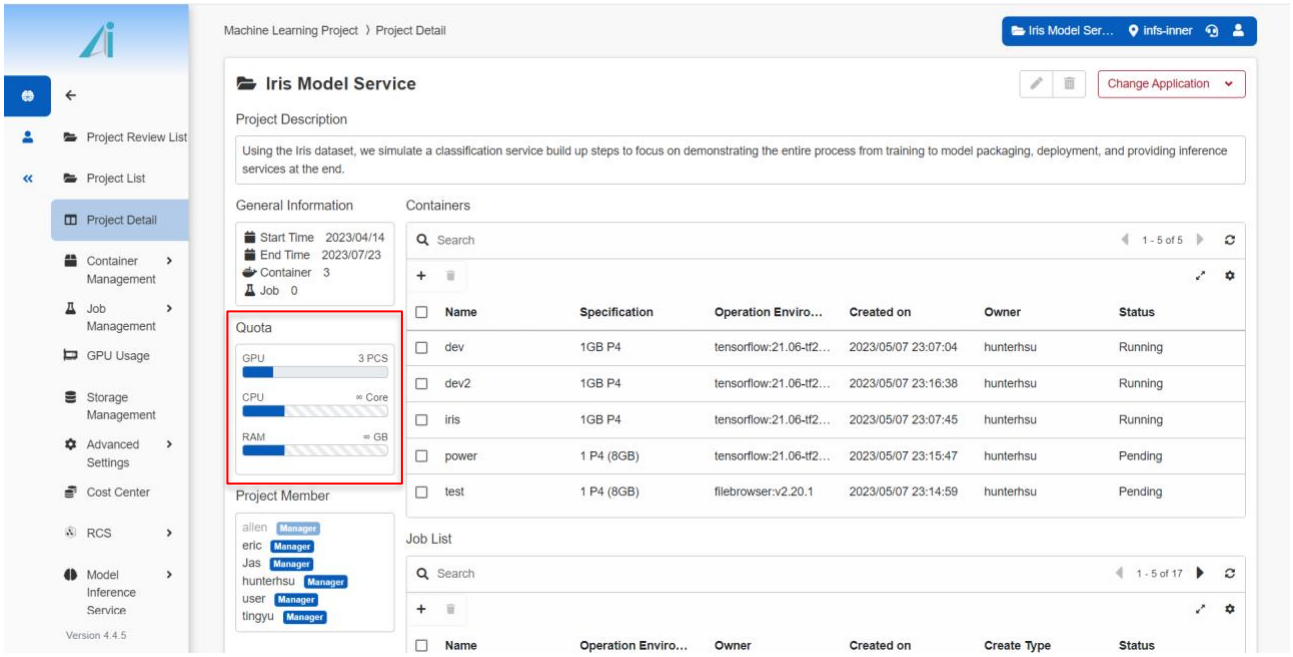


Once the project extension is approved. The icon for the start date will turn blue. Hov er over the icon and you can see the next start and end date for the project.



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.

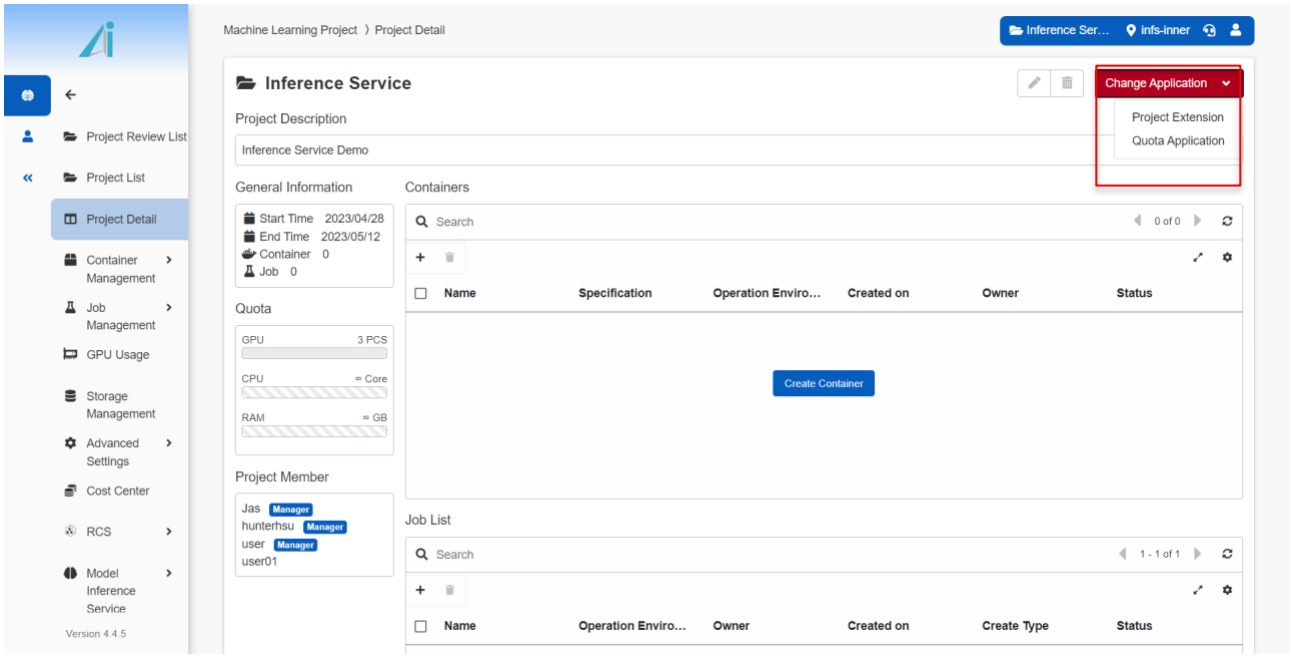


The area in the red box shown in the above figure records resource quota and availability for GPU, CPU and RAM quota of the project.

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

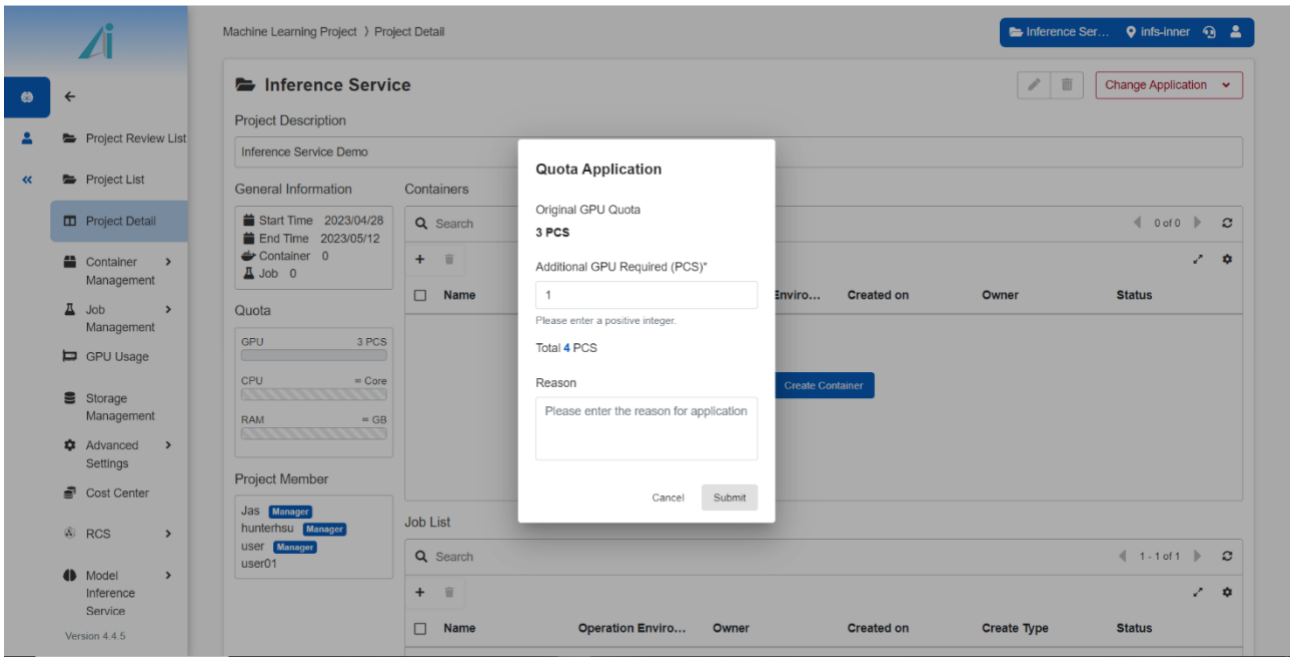
A. Project Change Application

In change applications, you are able to apply for a duration extension and quota inc rease, 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 applicati on 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

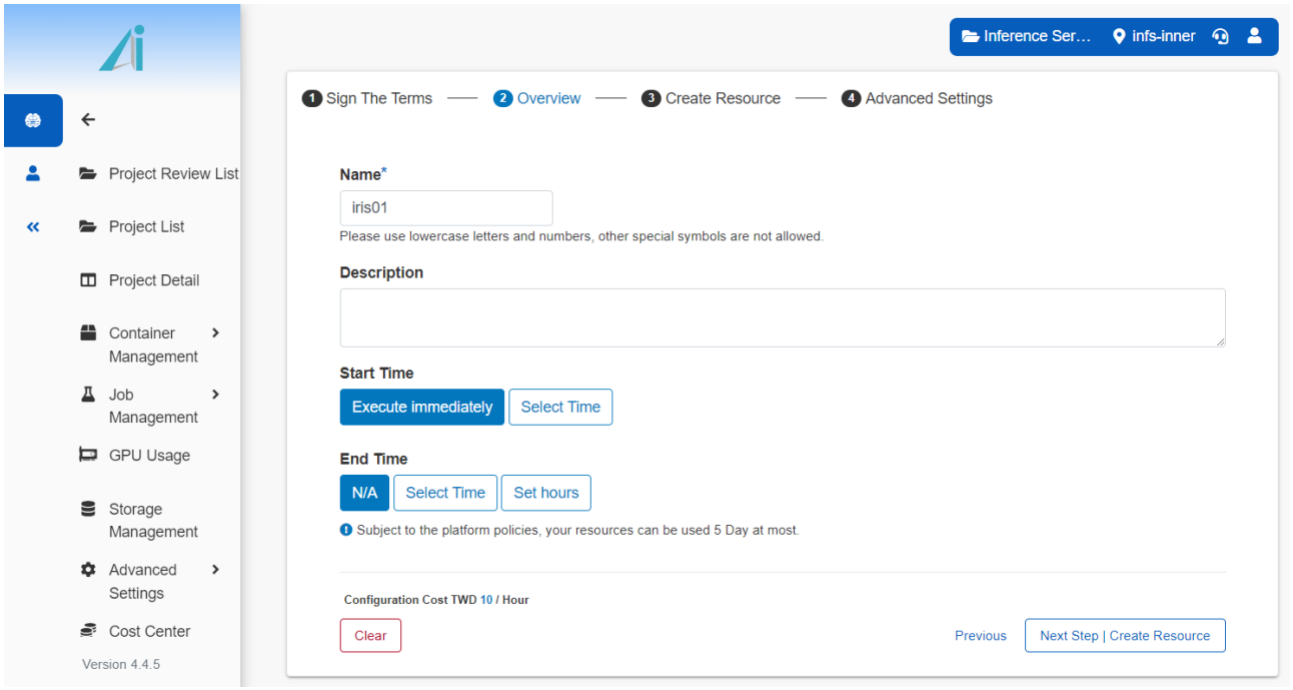
(1) Go to "Container List" and click on "+"

(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.

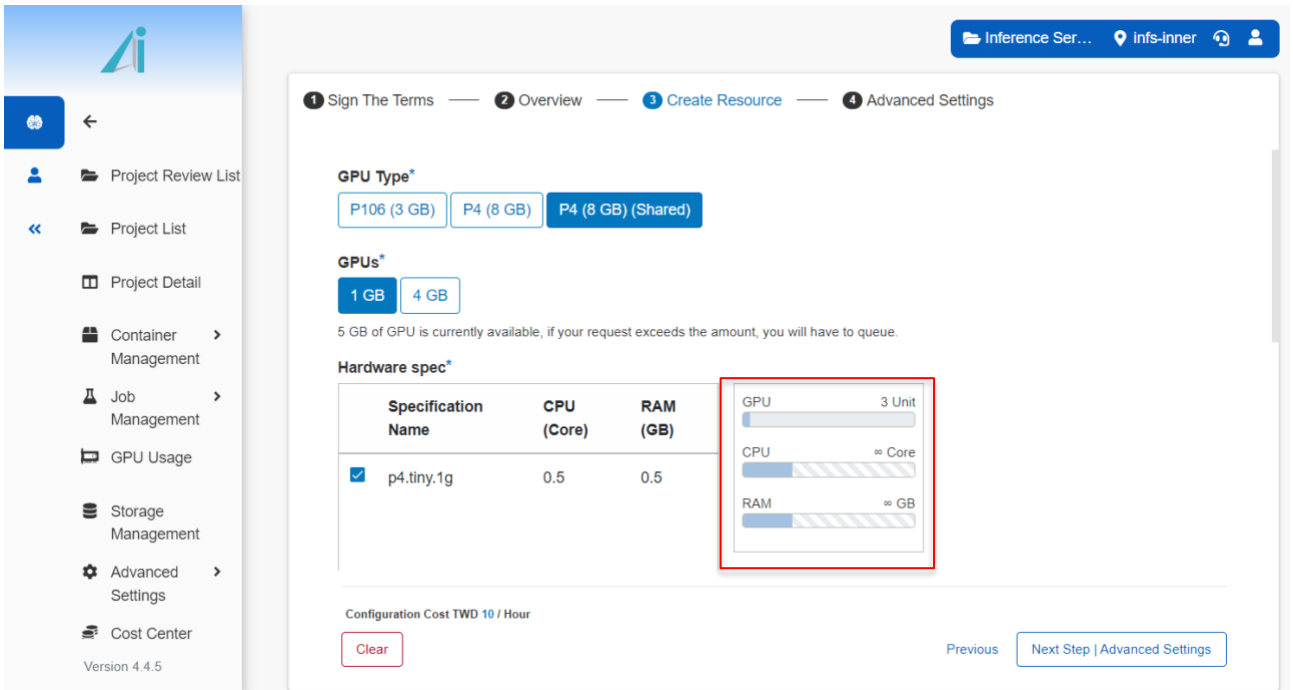
Click "Next Step | Create Resource" to continue



(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 m emory.

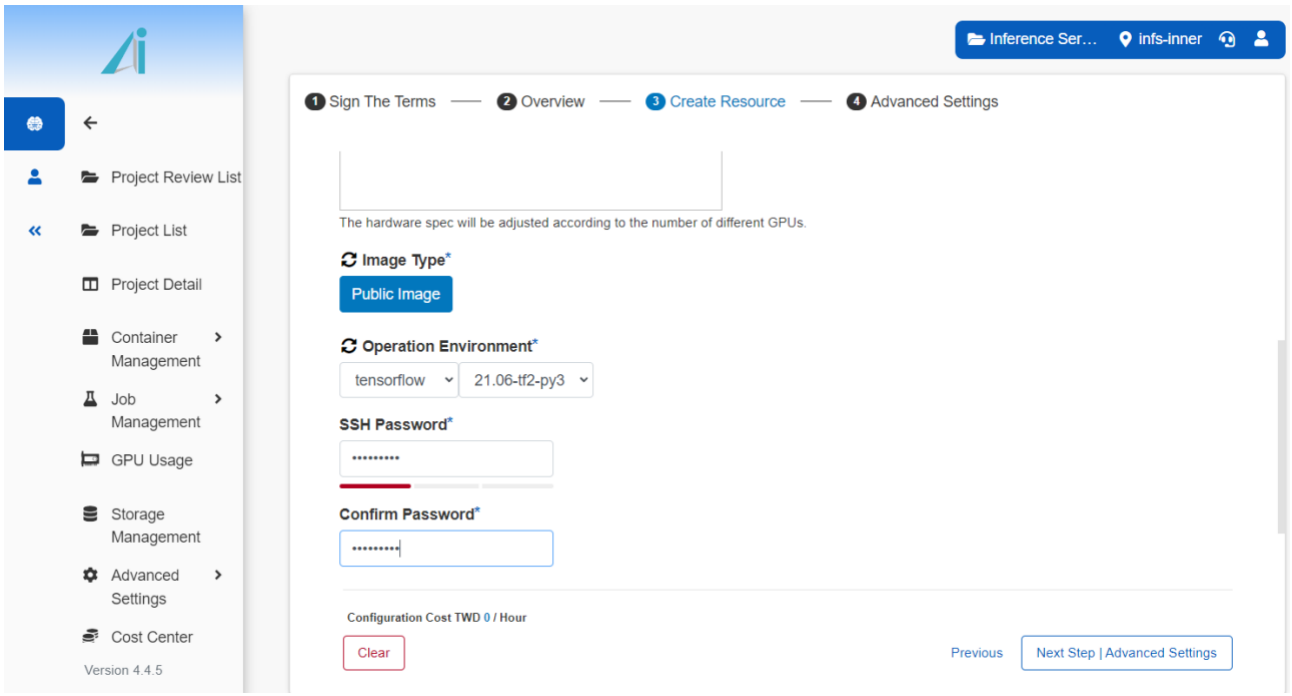
The red boxed area to the right of Hardware specifications shows the resource availabil ity information. It will change based on the selected specification.



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

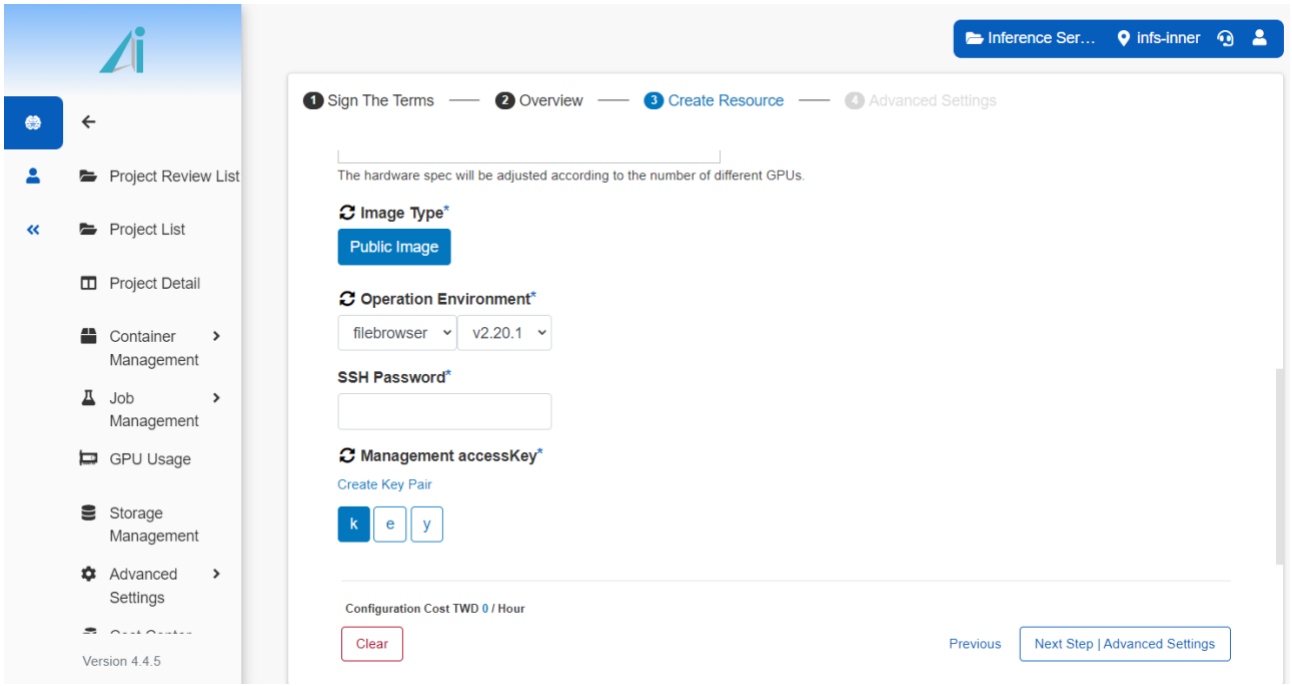
You have to create at least 1 custom image for the option to appear.

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



(8) Key pair management

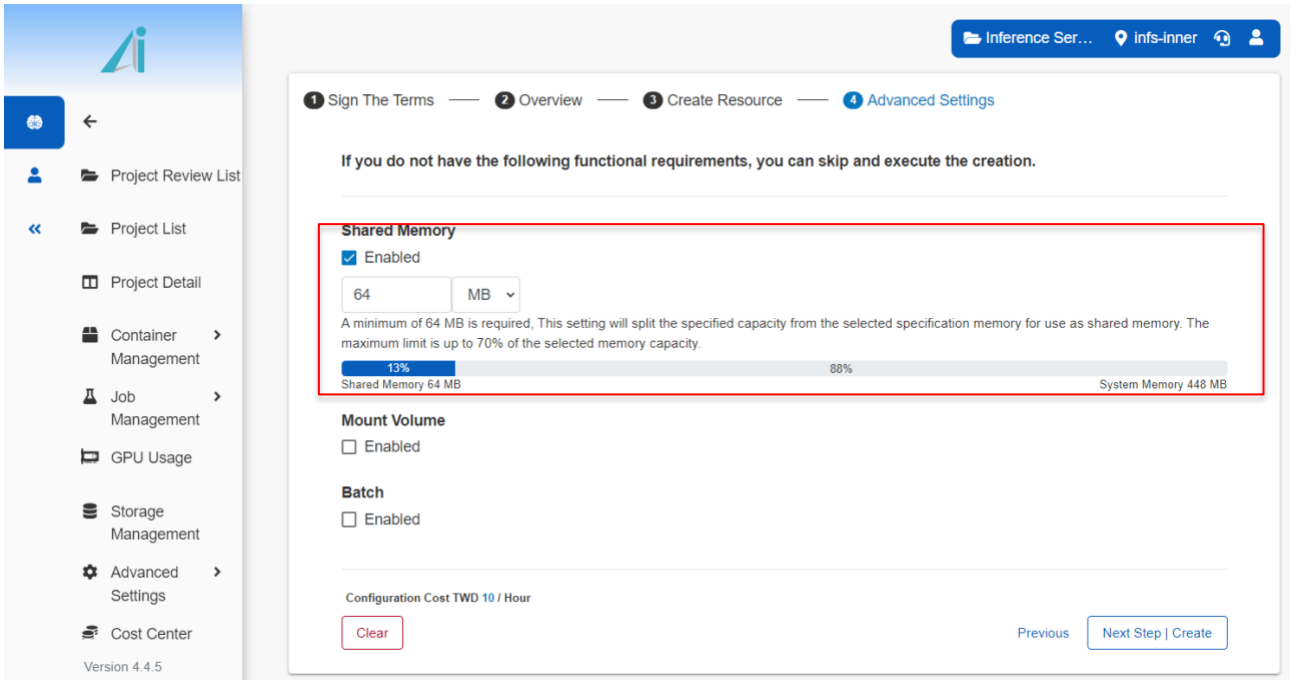
When a user has previously created a key on the platform and has enabled the SSH log in method with the selected image template, they can choose which key pair to use for the connection.



Click “Next Step | Advance Settings” to continue

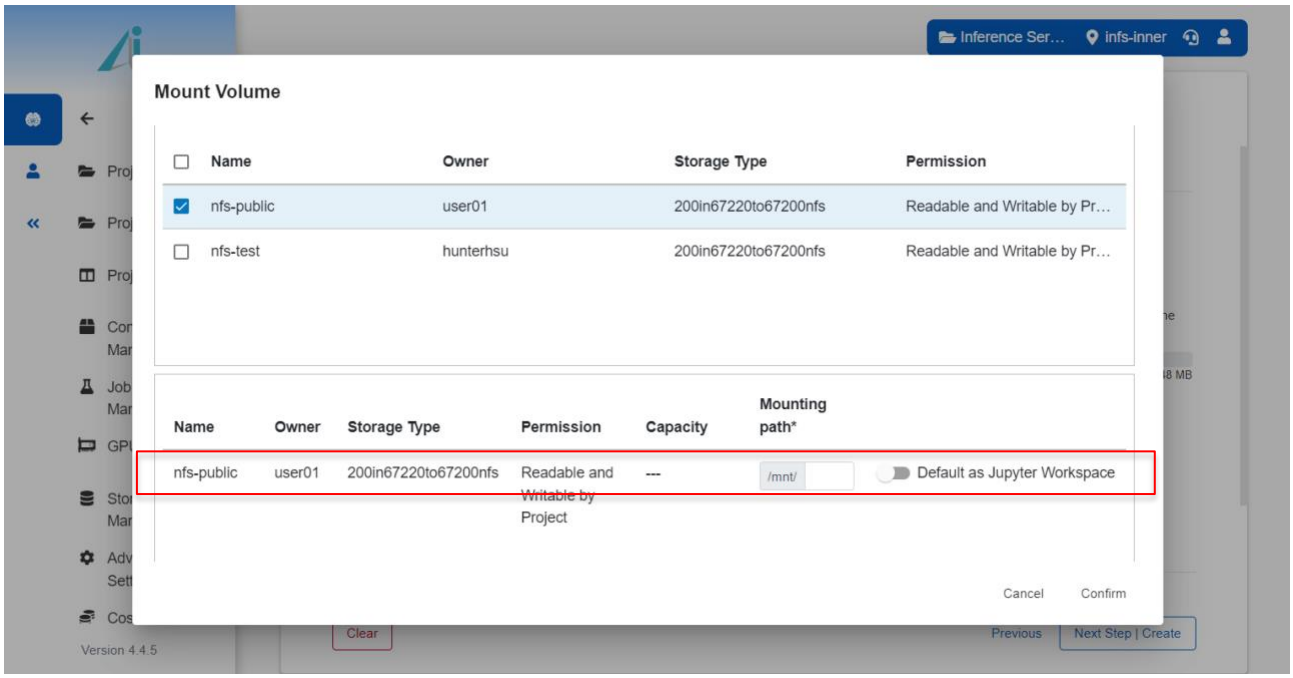
(9) Enable the option for "Shared Memory."

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...
infs-inner
Project Review List
Project List
Project Detail
Container Management
Job Management
GPU Usage
Storage Management
Advanced Settings
Cost Center
Version 4.4.5
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%
88%
Shared Memory 64 MB
System Memory 448 MB
Mount Volume
Enabled
Batch
Enabled
Configuration Cost TWD 10 / Hour
Clear
Previous
Next Step | Create

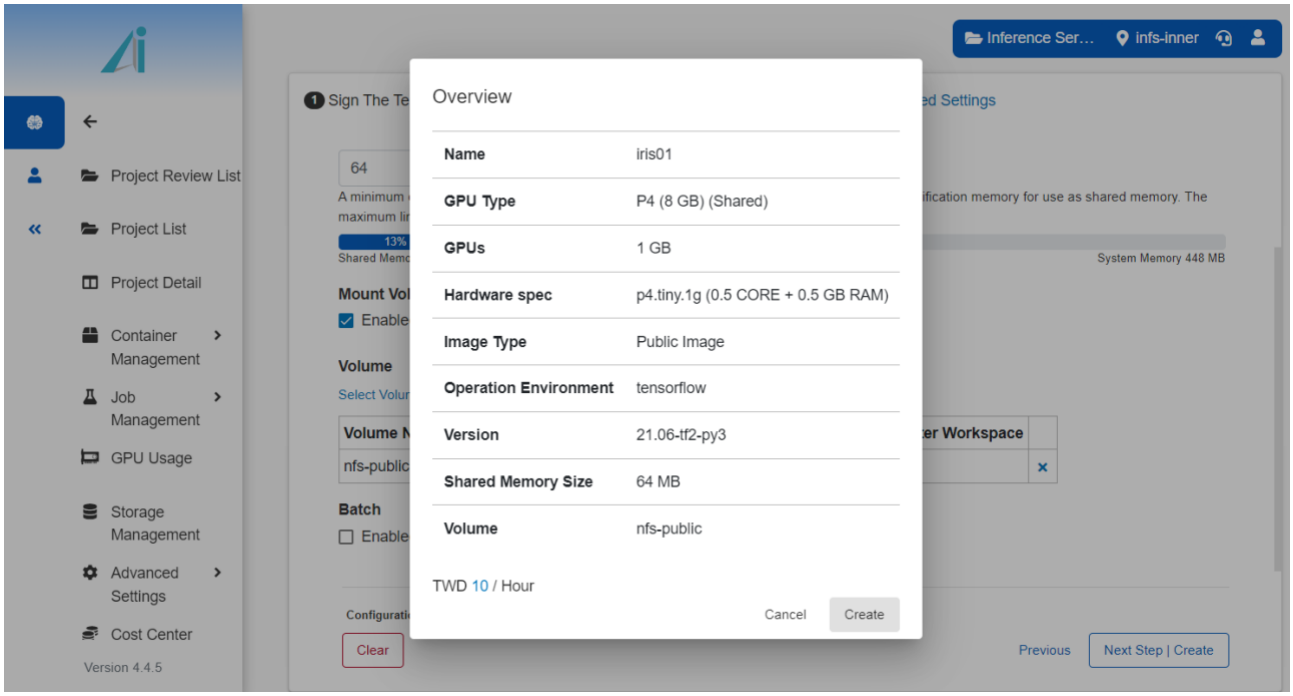
(10) Mounting Storage Device

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

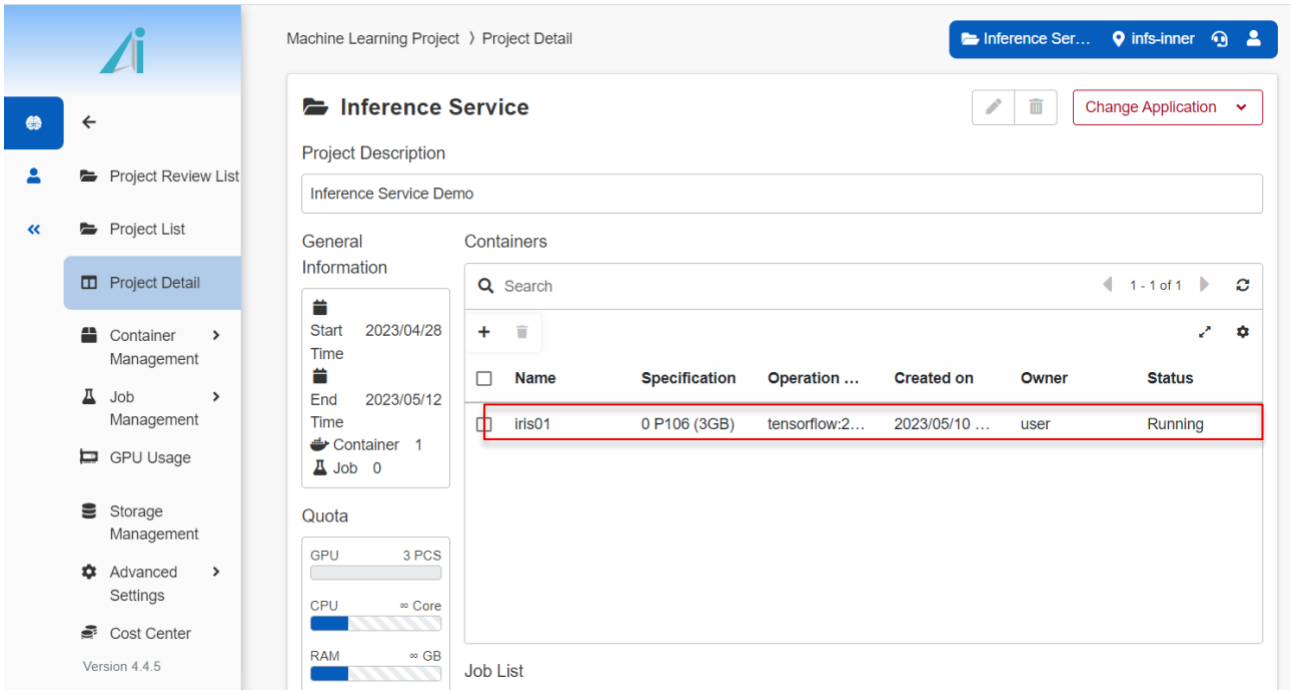


* Remark: you will have to create a storage to mount in storage management

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



(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:

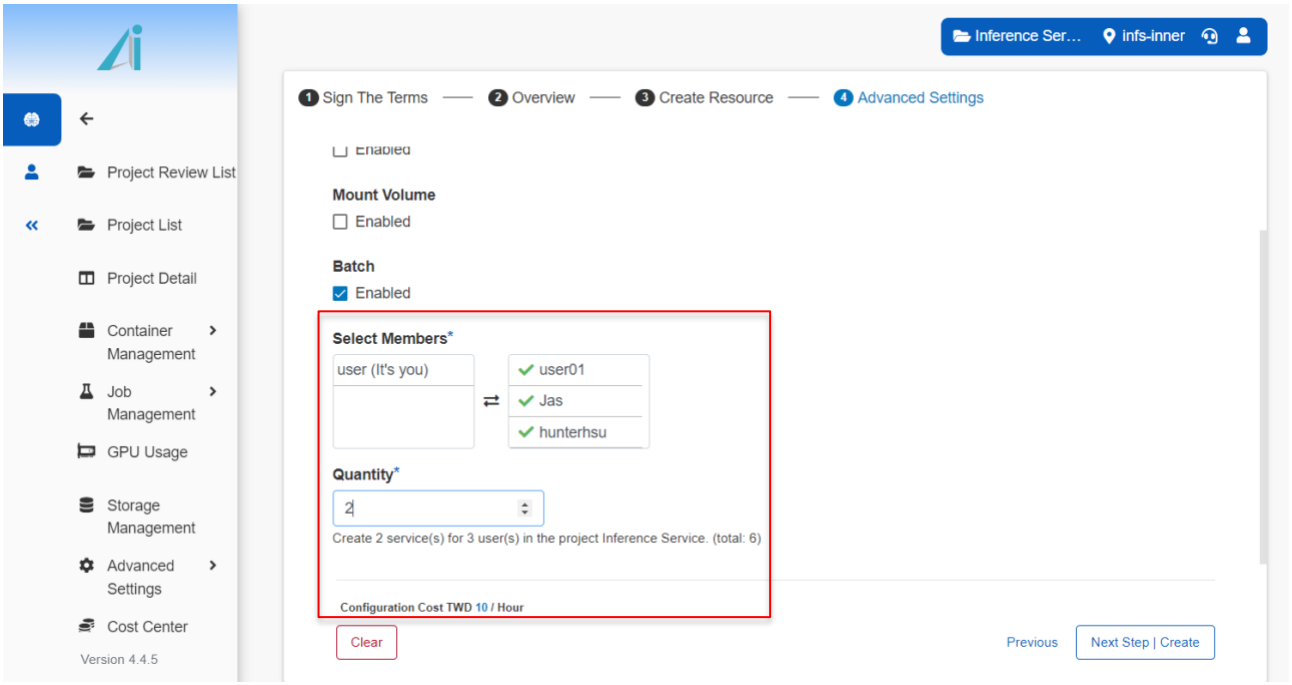
(1) Perform general container creation steps.

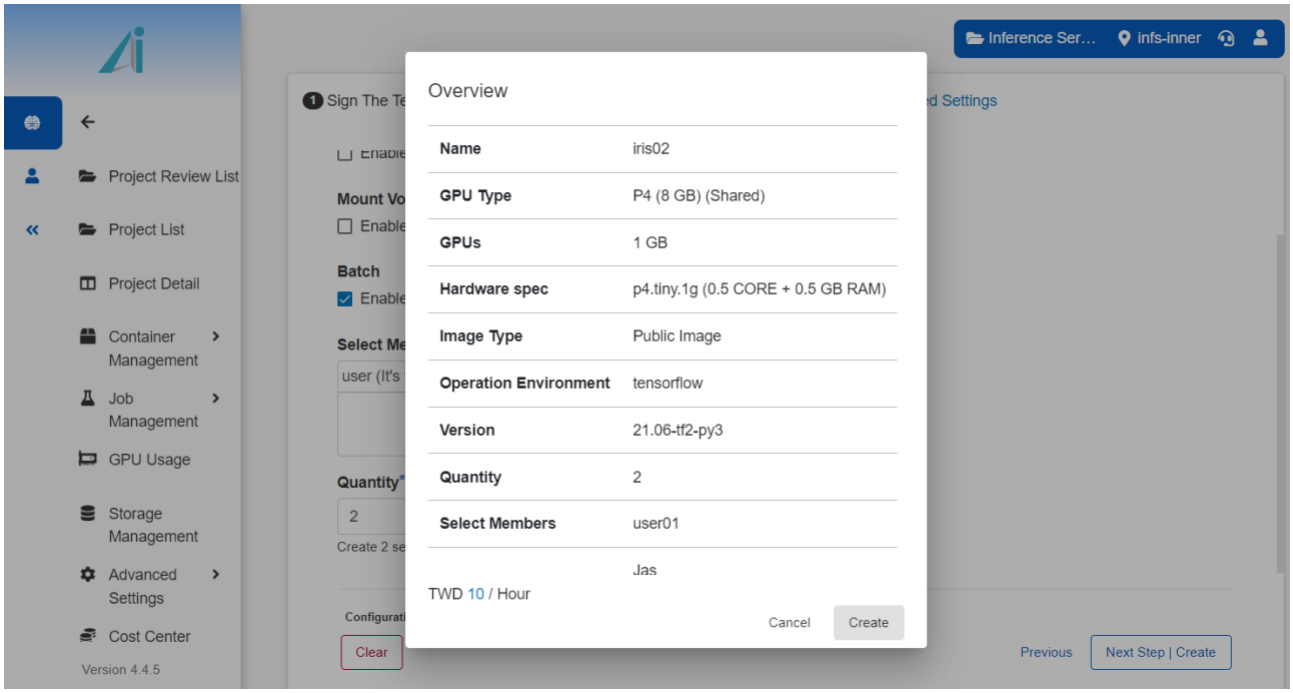
(2) Refer to the red-boxed area in the below figure, enable batch creation in advanced set tings by ticking the box.

(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.

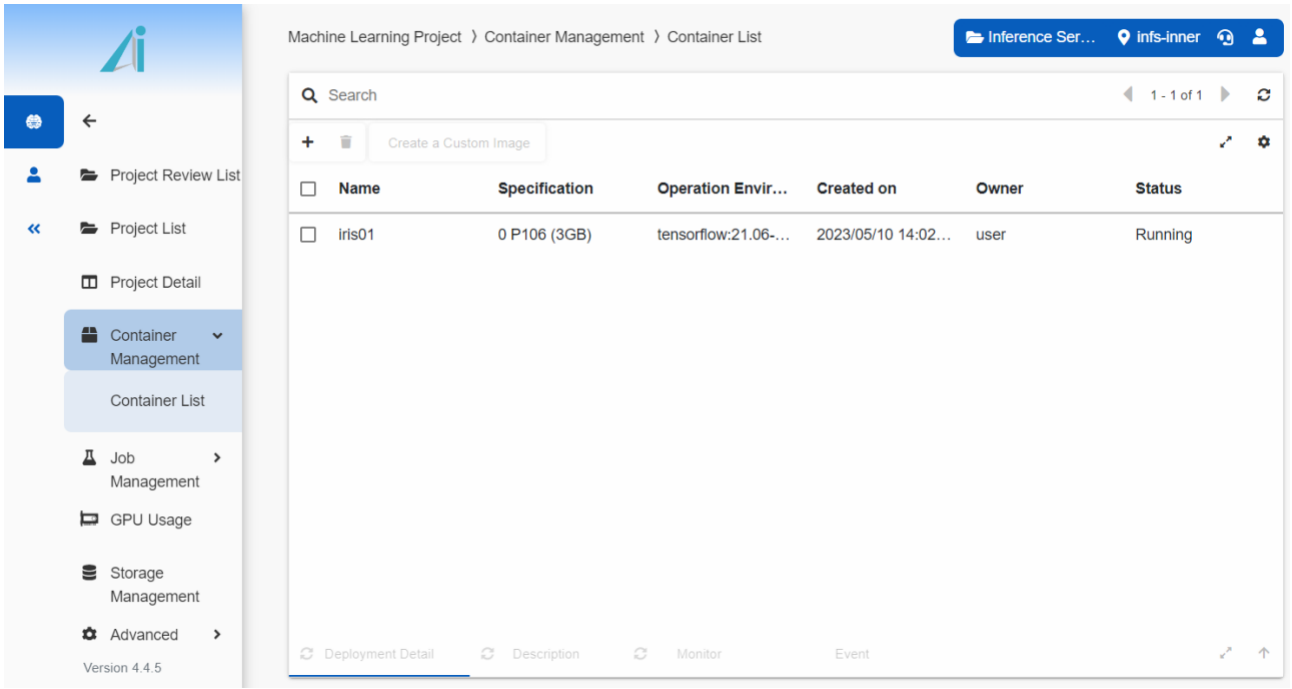
(5) Click "Next | Create", confirm the details in the overview, confirm the setting.



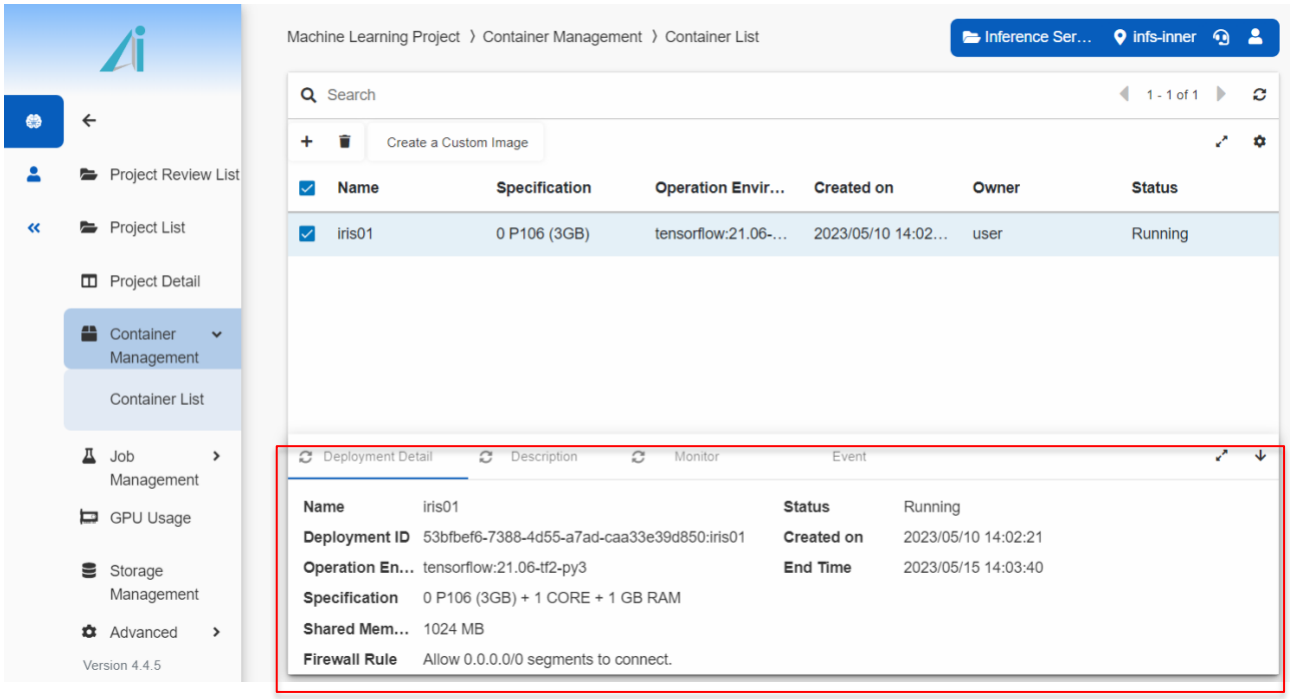

Inference Ser...
infs-inner
Project Review List
Project List
Project Detail
Container Management
Job Management
GPU Usage
Storage Management
Advanced Settings
Cost Center
Version 4.4.5
Overview
Name
iris02
GPU Type
P4 (8 GB) (Shared)
GPUs
1 GB
Hardware spec
p4.tiny.1g (0.5 CORE + 0.5 GB RAM)
Image Type
Public Image
Operation Environment
tensorflow
Version
21.06-tf2-py3
Quantity
2
Select Members
user01
Jas
TWD 10 / Hour
Cancel
Create
Clear
Previous
Next Step | Create

4.2 Container List

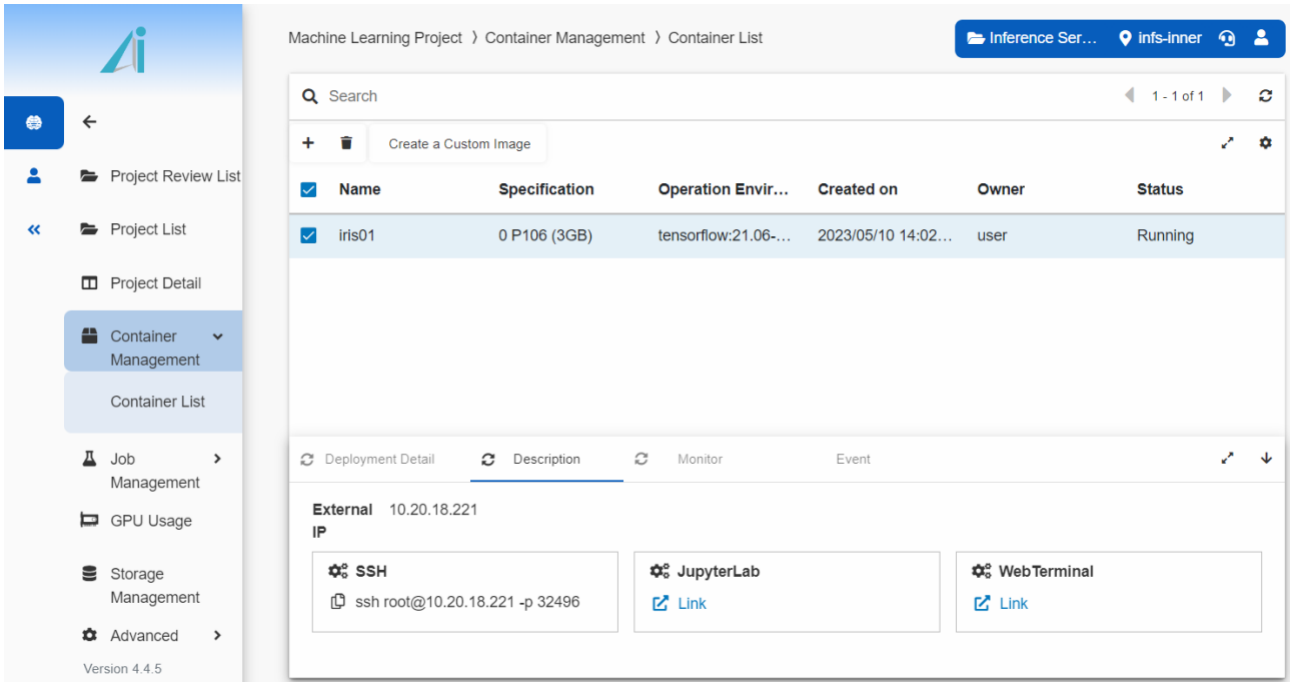
After the container creation is complete, you can go to the container list to view the list of created containers. The information on the list includes container names, specification (only displaying GPU information), operating environment, creation time, owner, and container status. When the status 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”, “descroptions”, “monitor” and “events”.


Machine Learning Project 〉Container Management 〉Container List
Inference Ser...
infs-inner
Search
1 - 1 of 1
Project Review List
Project List
Project Detail
Container Management
Container List
Job Management
GPU Usage
Storage Management
Advanced
Version 4.4.5
Create a Custom Image
Name
Specification
Operation Envir...
Created on
Owner
Status
iris01
0 P106 (3GB)
tensorflow:21.06-...
2023/05/10 14:02...
user
Running
Deployment Detail
Description
Monitor
Event
Name iris01
Deployment ID 53bfbef6-7388-4d55-a7ad-caa33e39d850:iris01
Operation En... tensorflow:21.06-tf2-py3
Specification 0 P106 (3GB) + 1 CORE + 1 GB RAM
Shared Mem... 1024 MB
Firewall Rule Allow 0.0.0.0/0 segments to connect.
Status Running
Created on 2023/05/10 14:02:21
End Time 2023/05/15 14:03:40

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

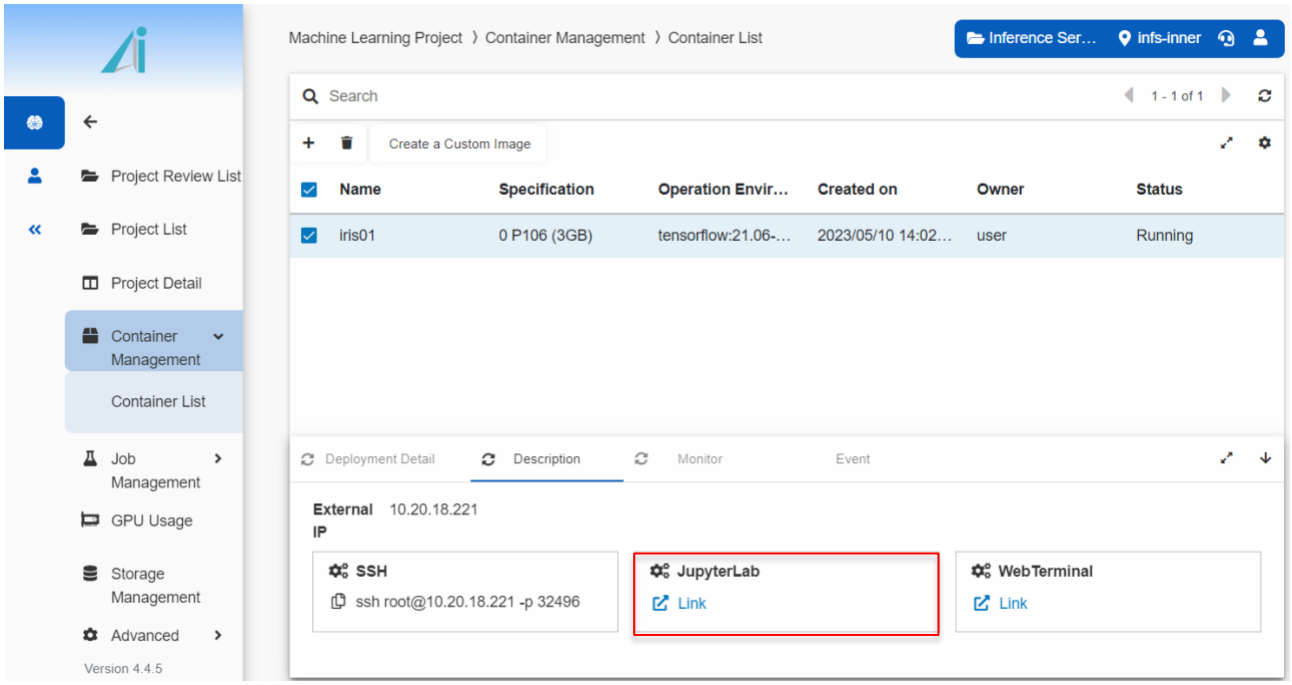
(1) SSH

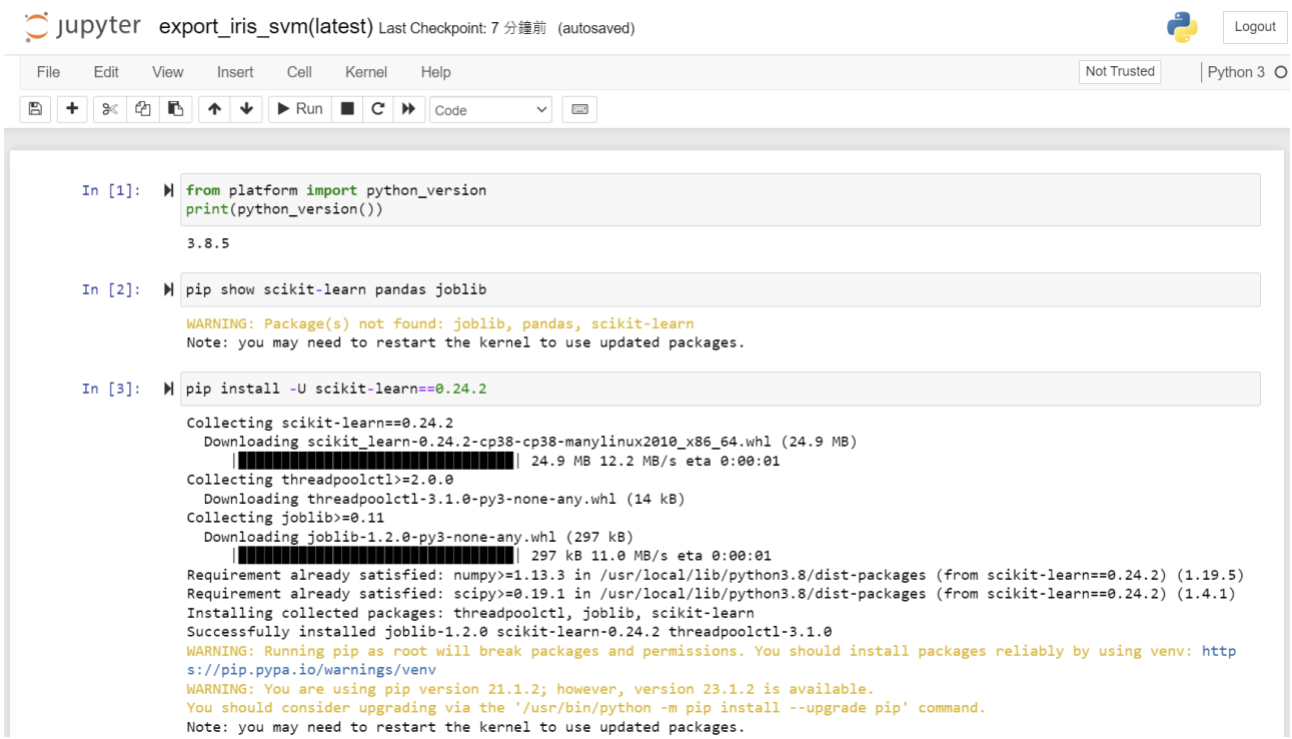
Using an SSH tool (such as PuTTY) in conjunction with the IP address, service port number, and the password entered during creation or the selected key, you can SSH into the container remotely for operations. Taking the example of viewing detailed information of the "dev" container, the SSH command would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

root@dev-5b9559b4b4-tplvq
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

(2) Jupyter

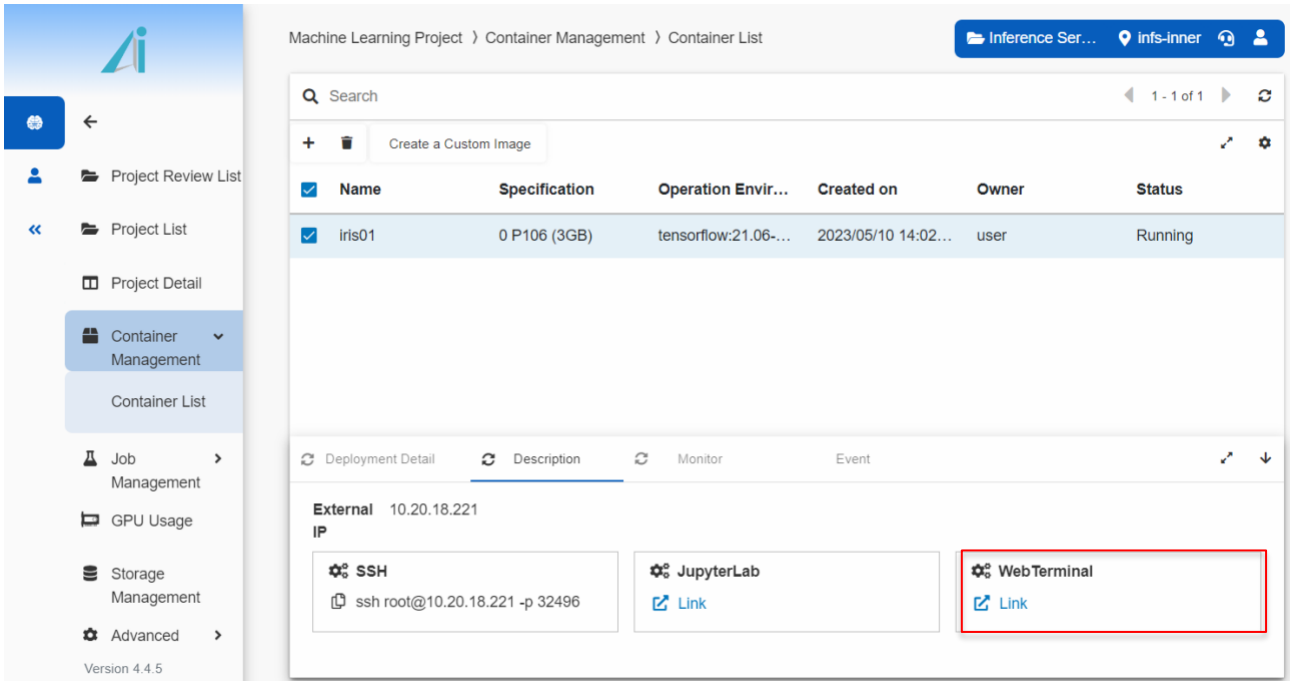
On the "Service Information" tab, clicking the "Jupyter Link" will directly go to Jupyter Notebook. Users can use it to write Python programs.

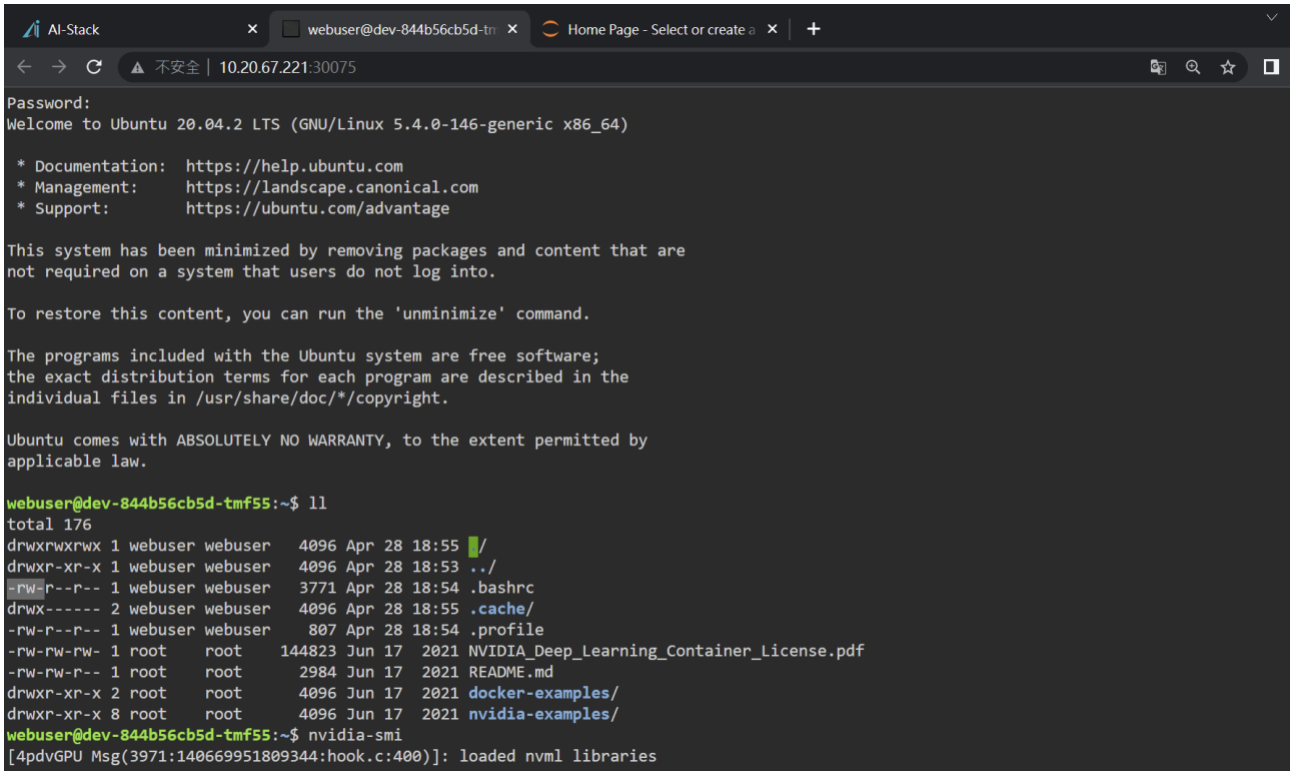




(3) WebTerminal

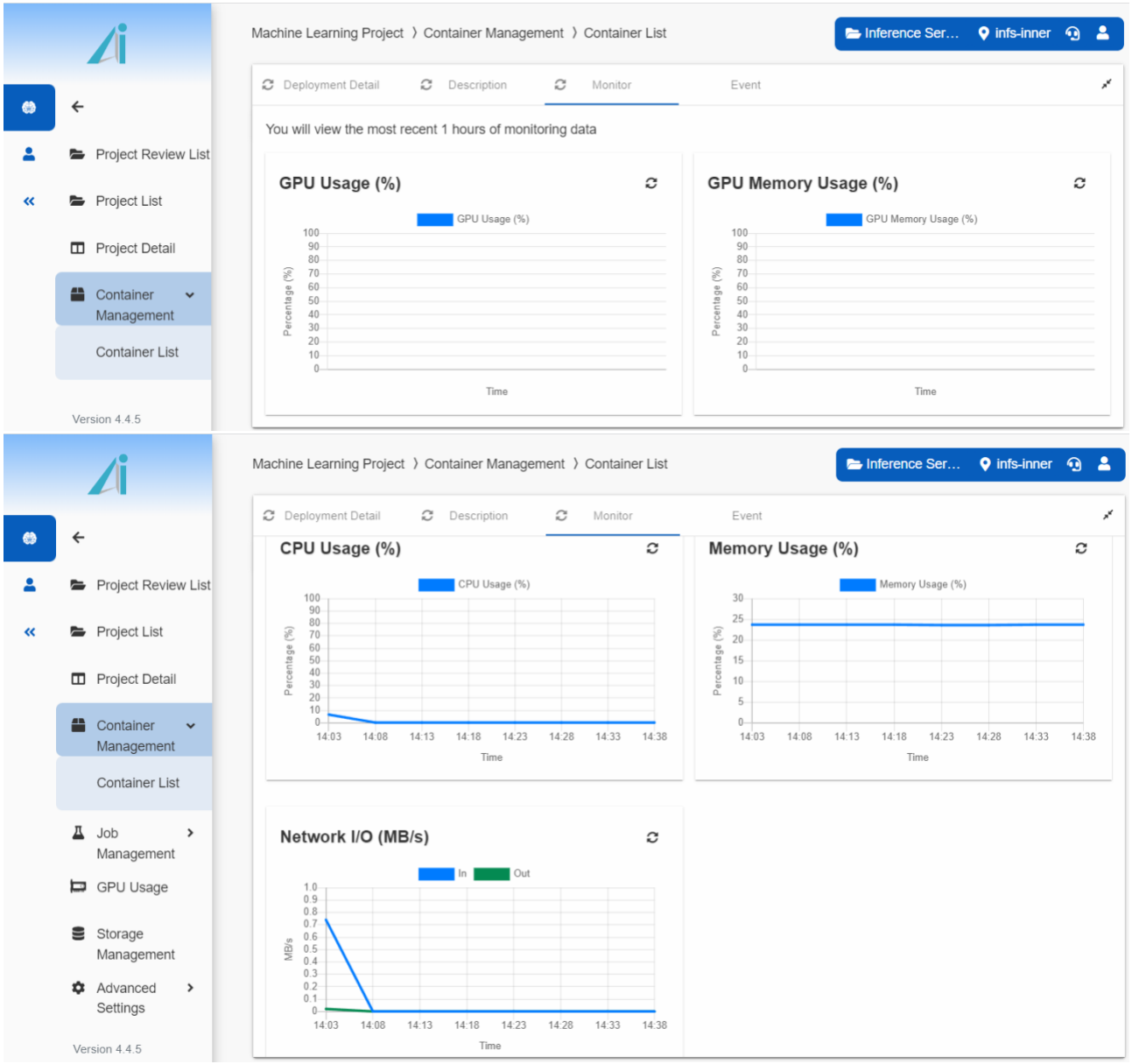
On the "Service Information" tab, clicking the " WebTerminal " will directly go to WebTerminal.





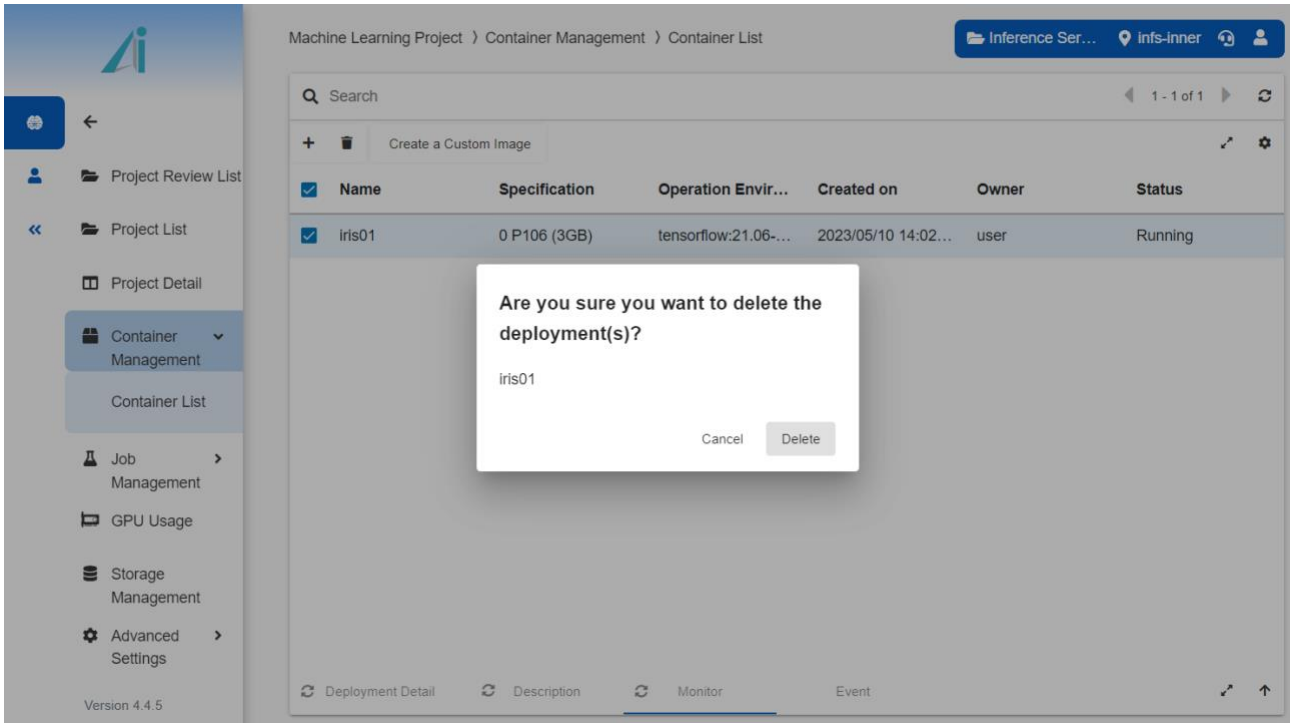
Provisioning

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].

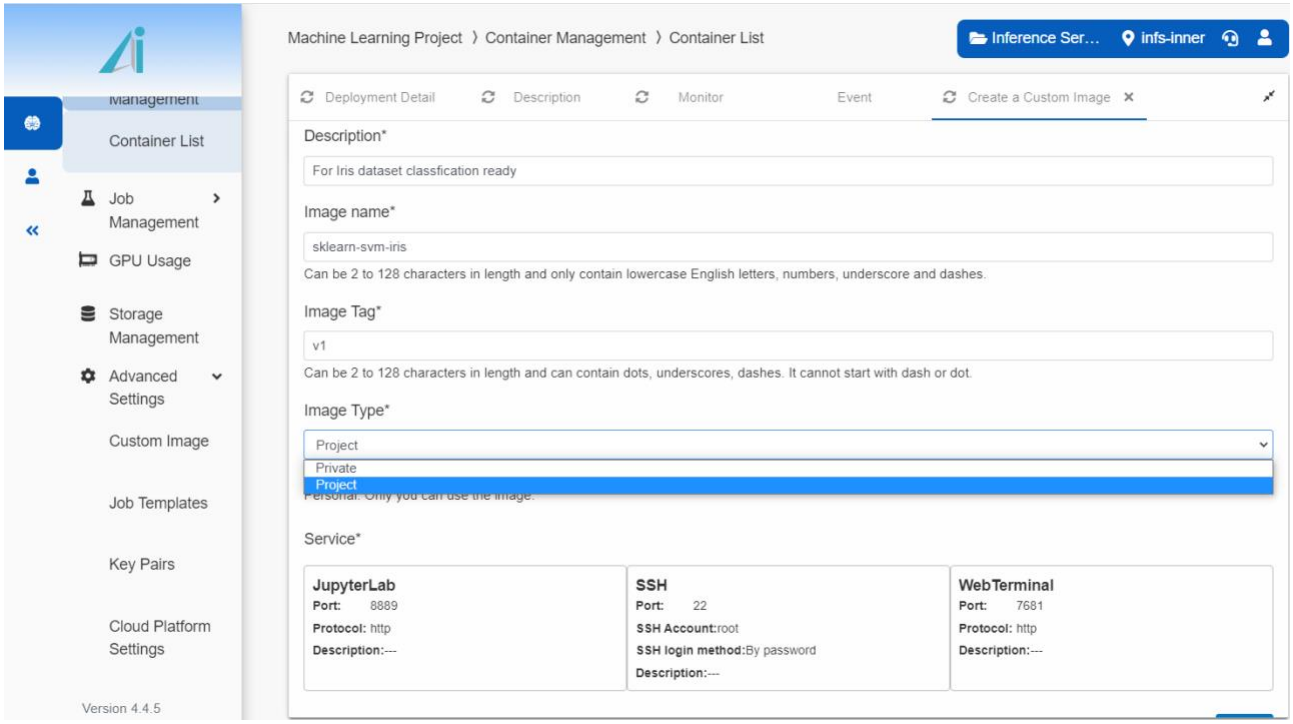


Note: Once a container is deleted, it cannot be recovered. Please ensure that you have backed up the files and data or have stored the data that needs to be retained in a separate location under the mounted device path to ensure that the data can be retrieved later.

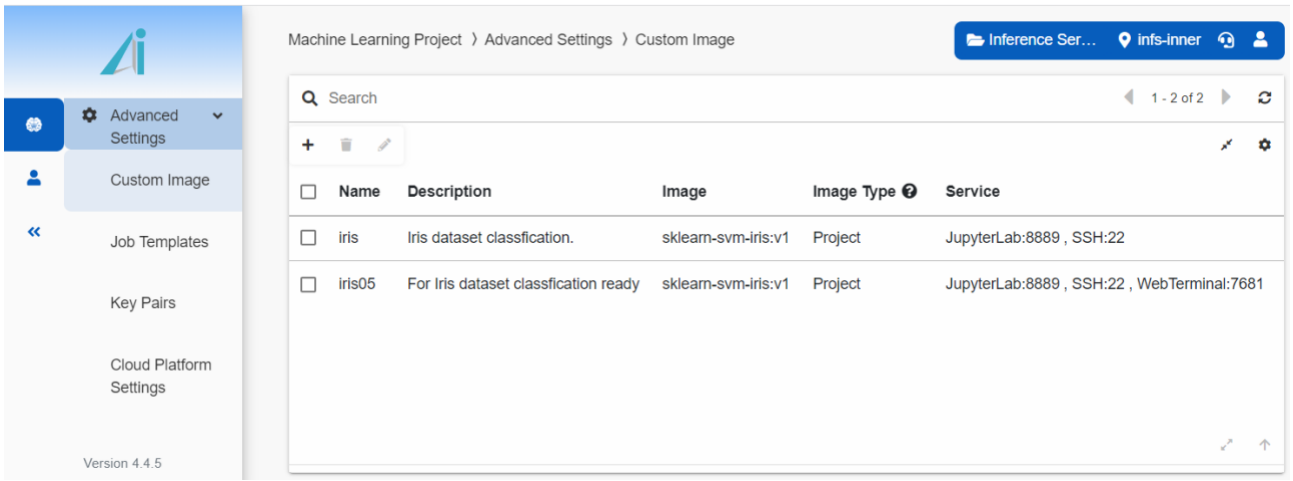
4.4 Create Custom Image

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 ori ginal 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 pr oject members. Refer to Section 8.1 for further details,



You can view your custom images under “Advanced Settings > Custom Image” pag e.

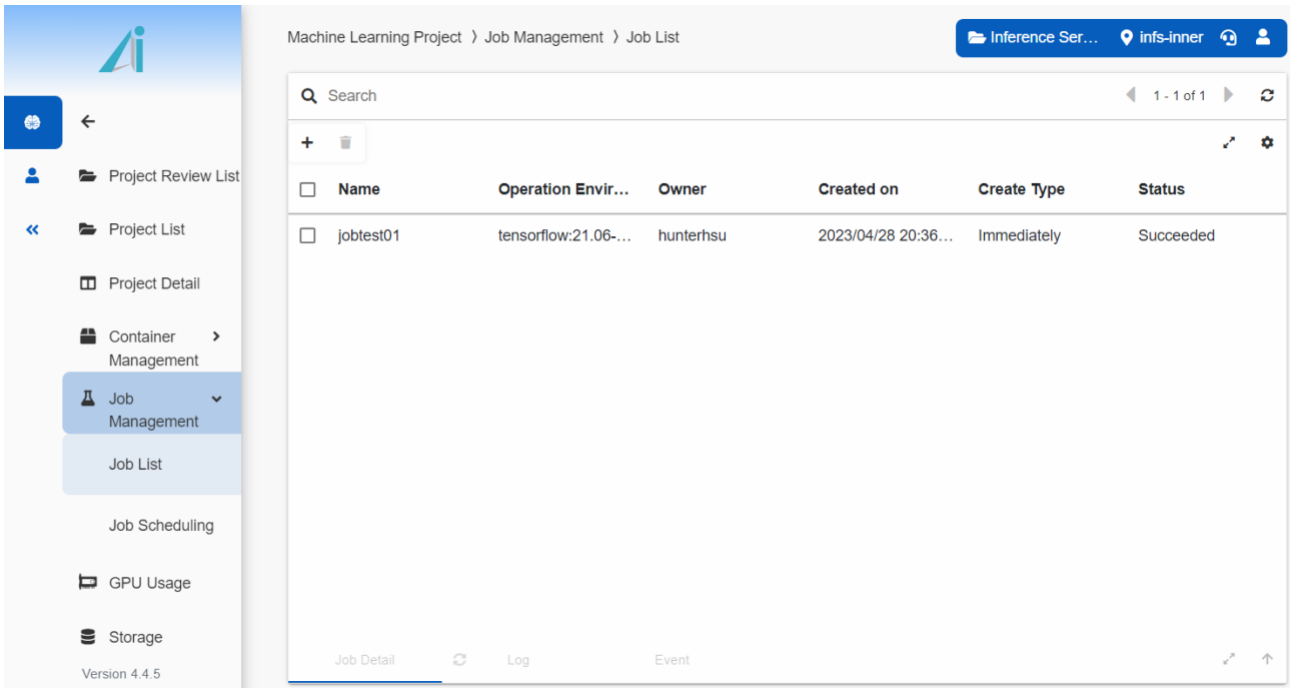


5. Job Management

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.



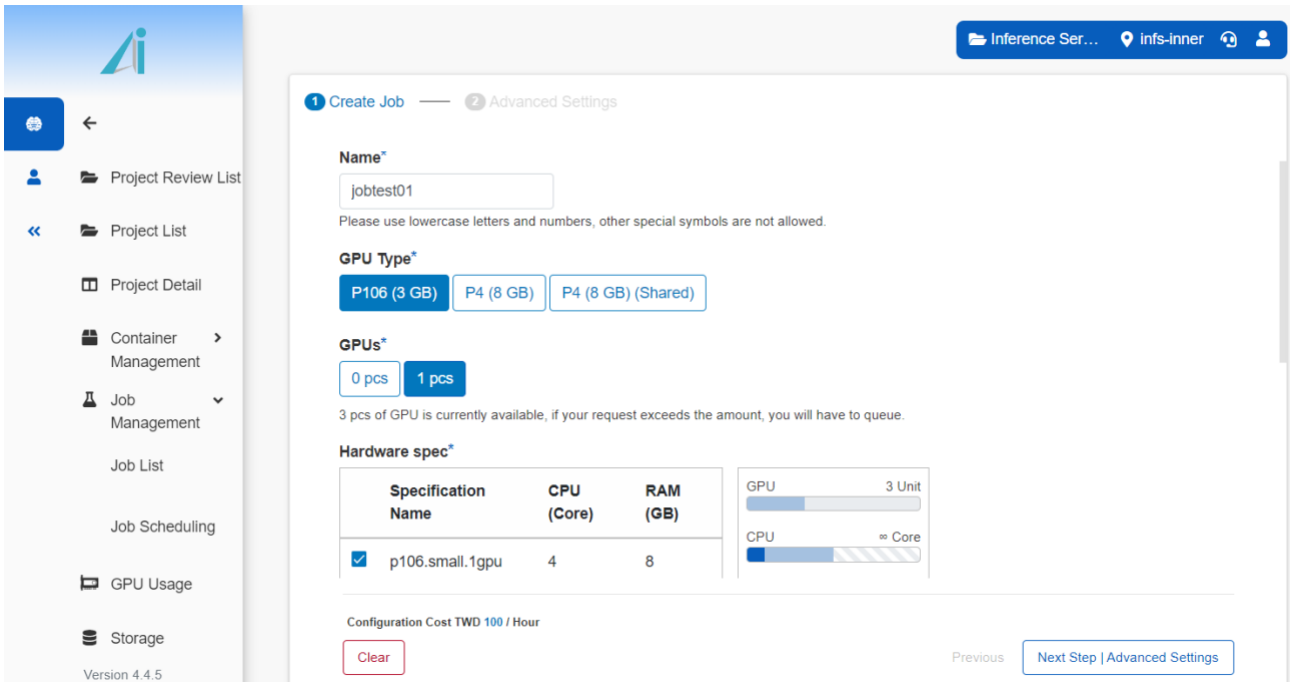
Create Jobs

Steps for job creation:

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

(2) Enter the name of your job.

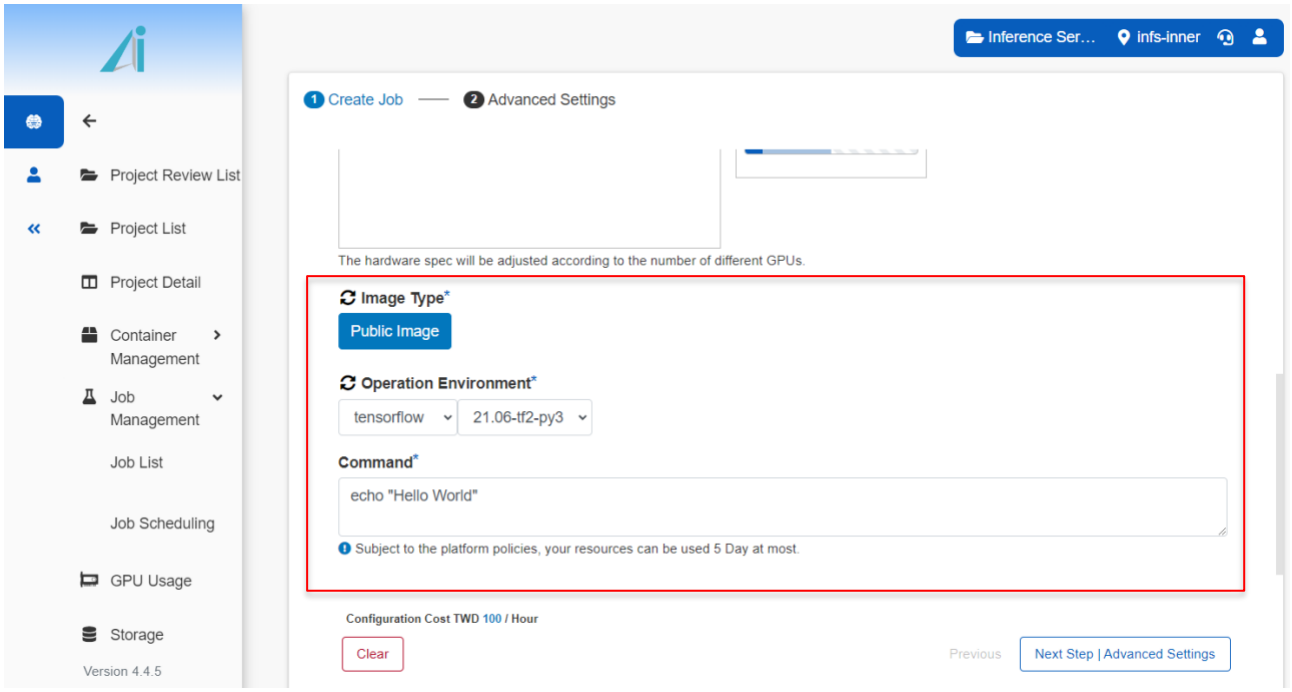
(3) Select the preferred GPU type, GPU quantity and the hardware specifications (in cluding CPU cores and memory)



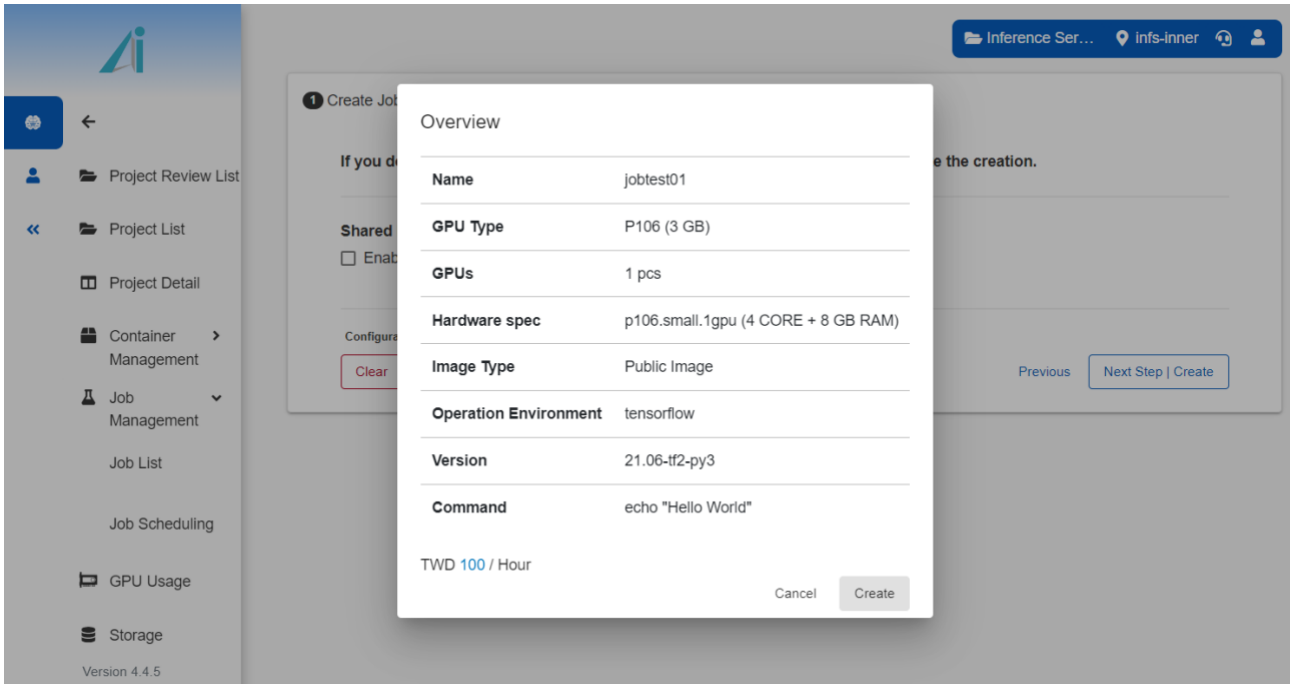
(4) Select either a public image or a customer image (if a customer image is prebuil t)

(5) Select the framework under operation environment.

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

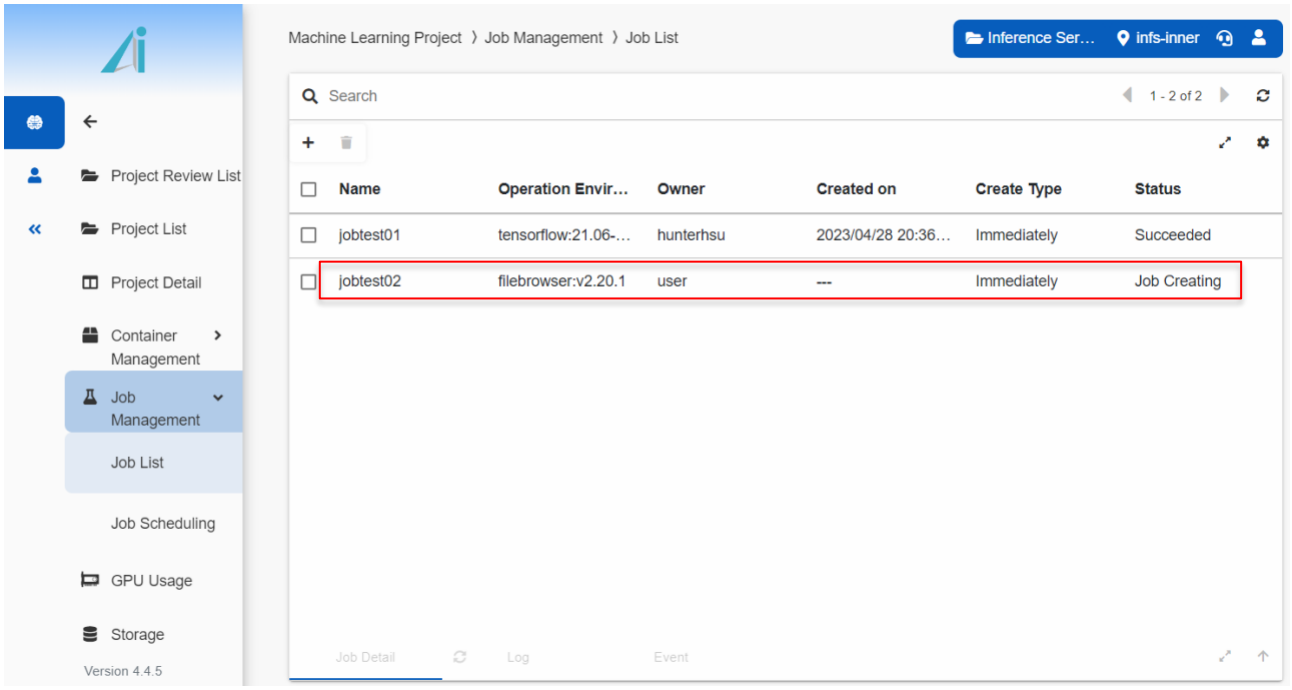


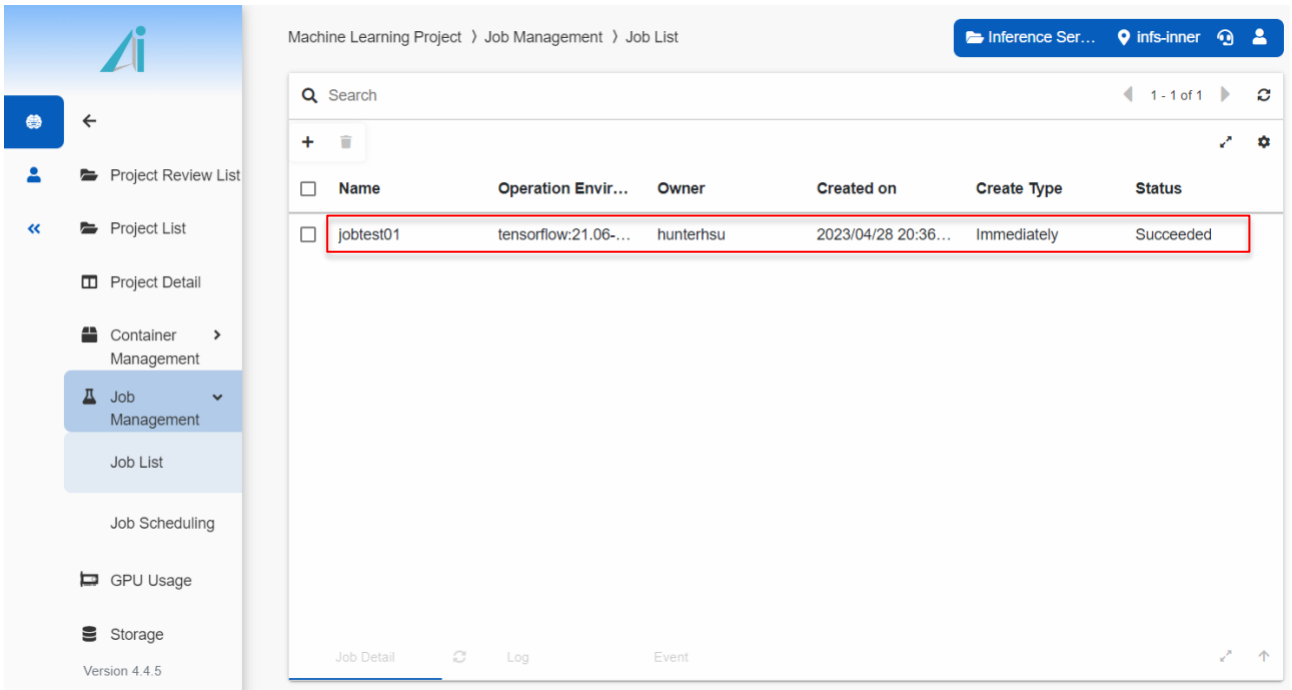
(7) Create a job by clicking on next.



(8) Confirm the details in the overview.

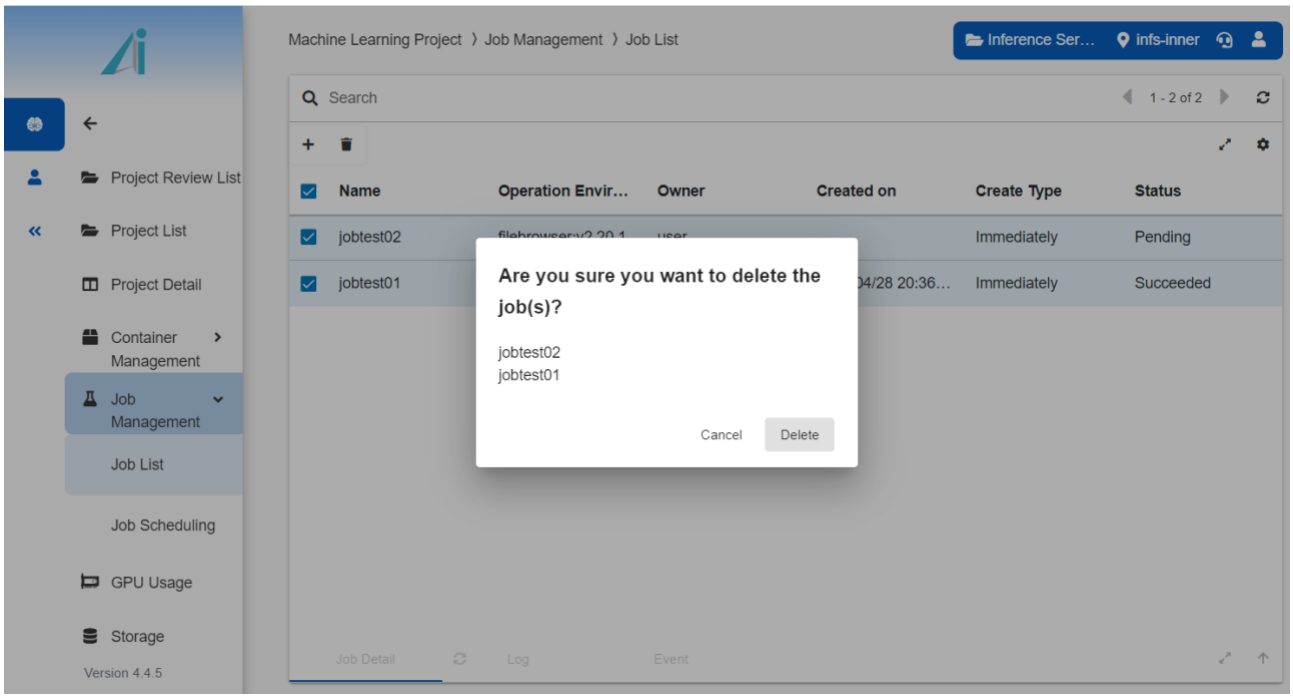
(9) You can view the progress in the job list

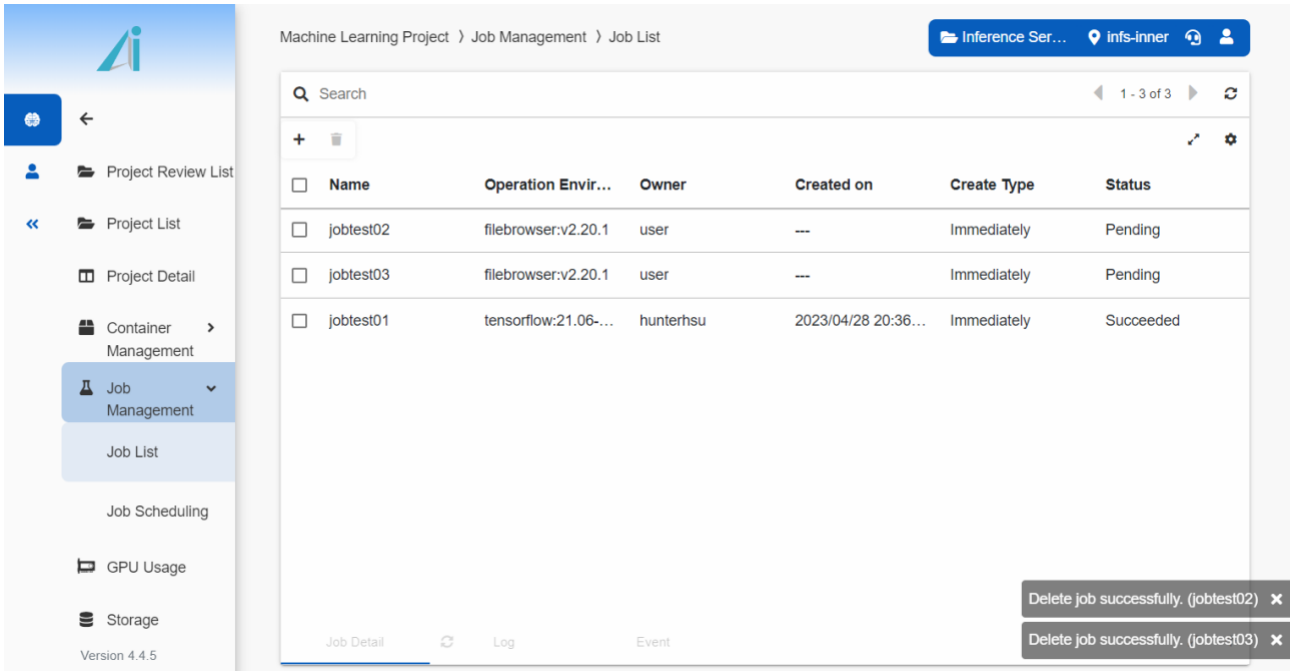




Delete Jobs

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。

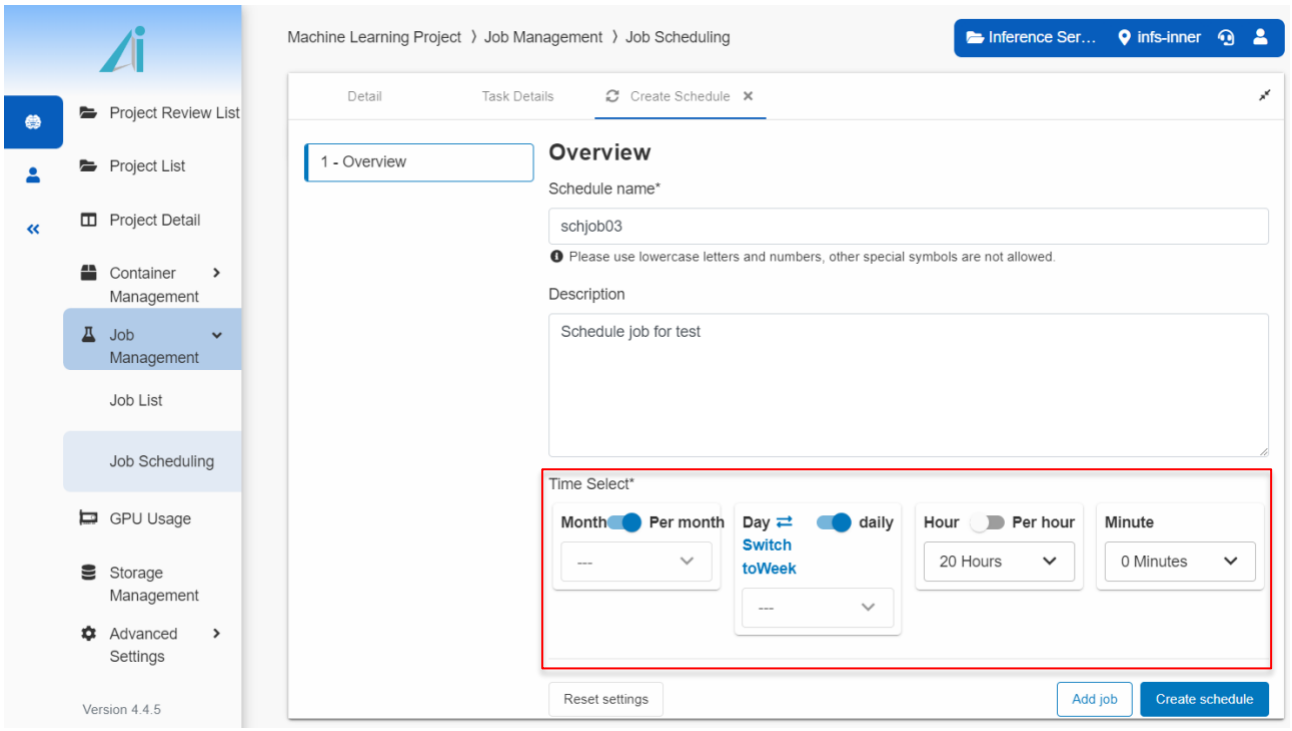
a. Go to Job Management > job schedule

b. Click [+] and you will be directed to the page to create new schedule

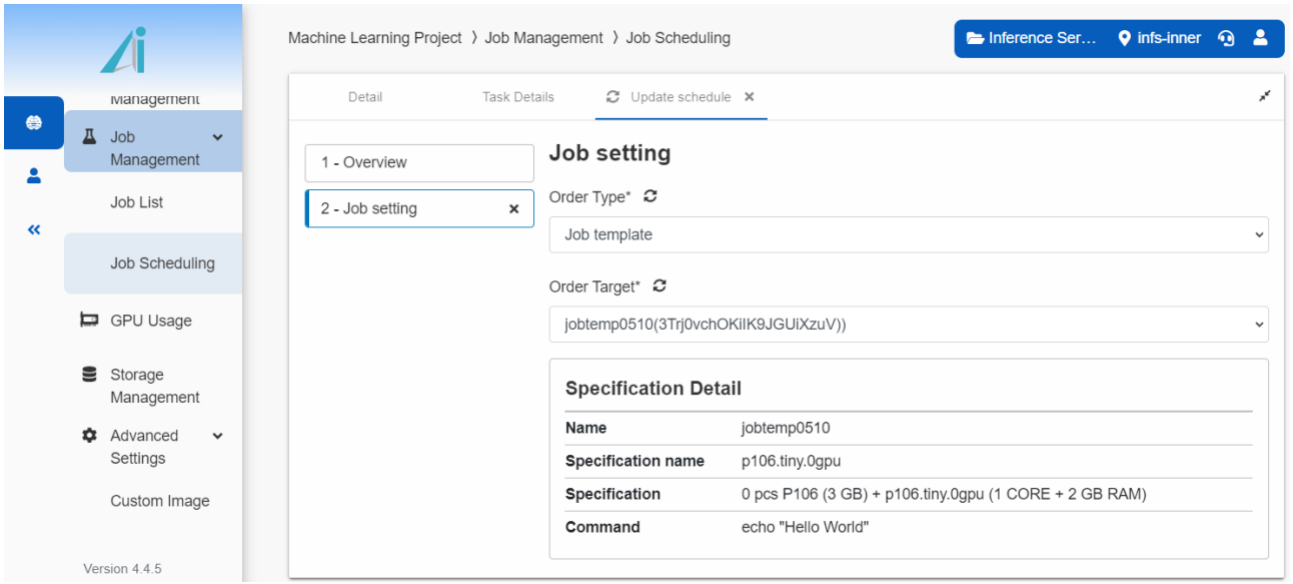
c. Enter the name and the description of the schedule

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).

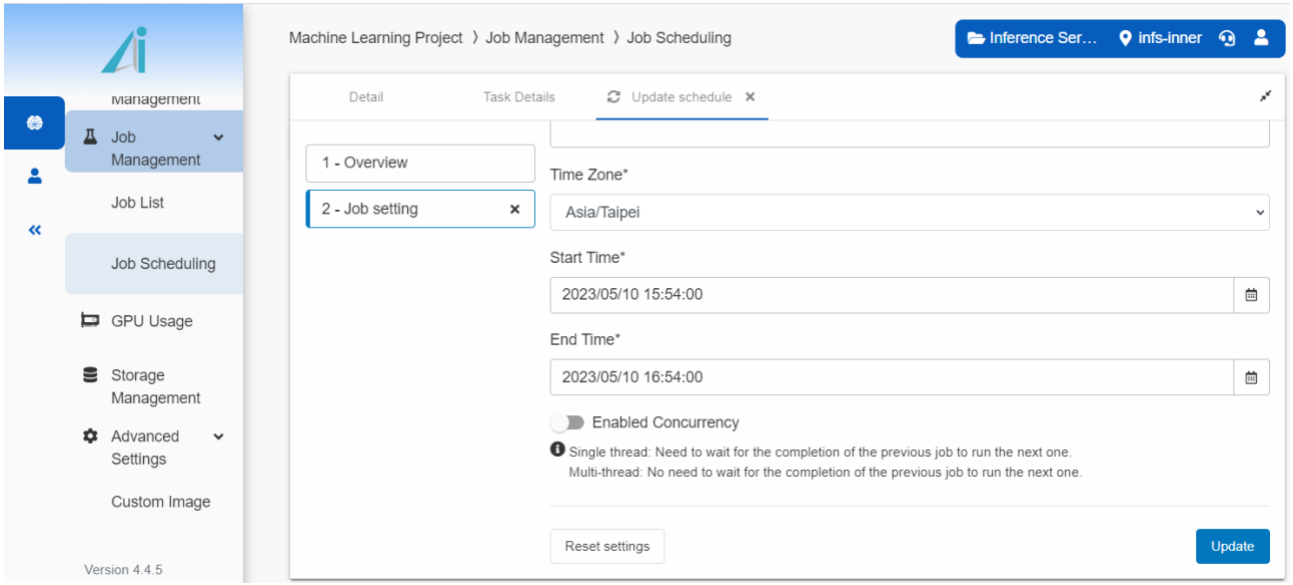


g. Select the time zone, start time and end time

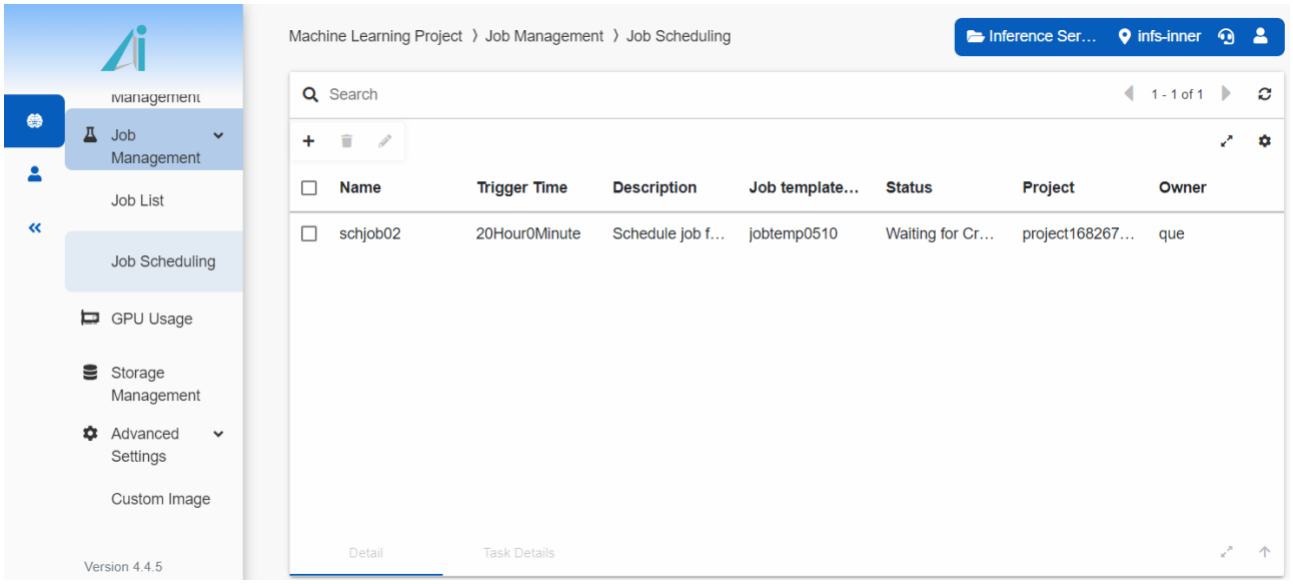
h. Confirm whether to activate multi-thread.

- single thread: Jobs are executed by sequence
- multi thread: Jobs are executed in parallel.

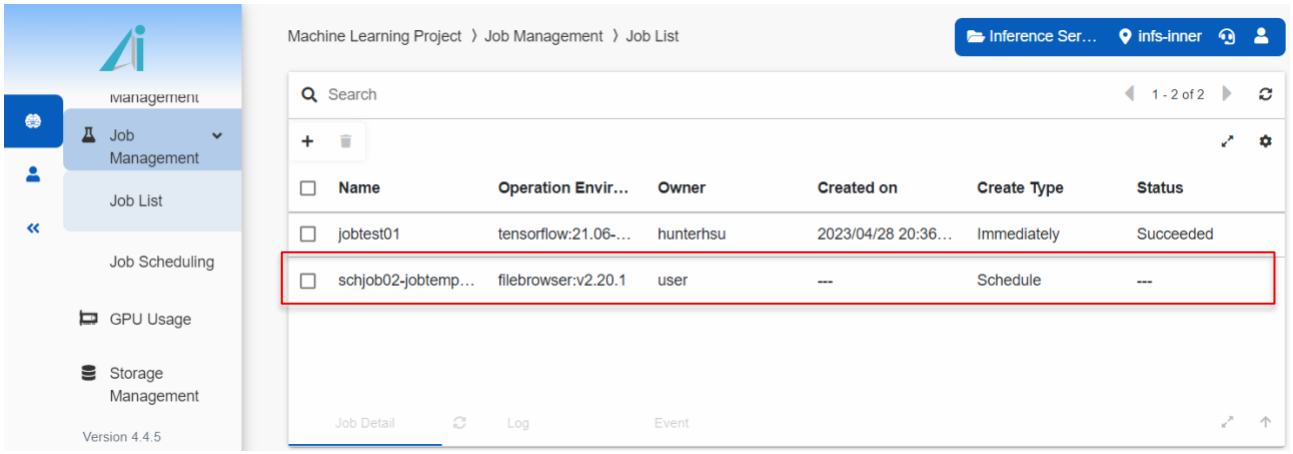
i. Click on the create button to confirm job creation.



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.

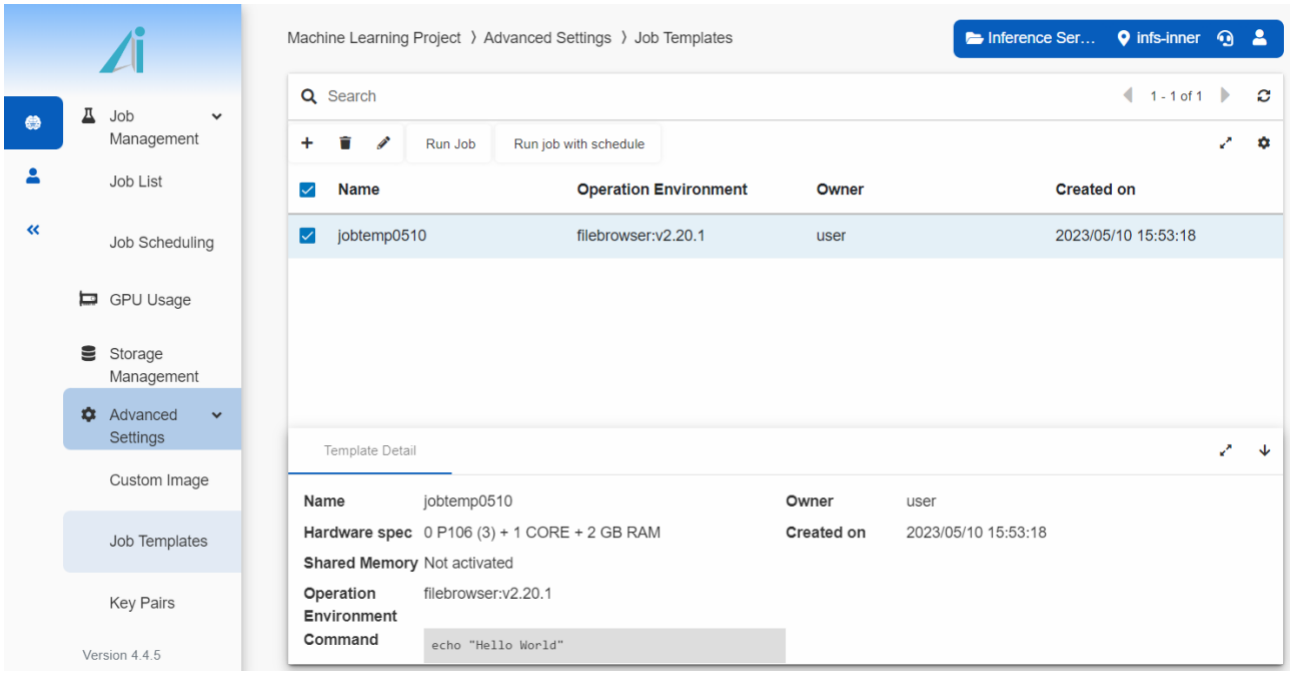


Note: You will have to delete the job schedule to delete it.

(2) From the "Advanced Settings > Job Template" page, select the for the job template to be created or associated with a job schedule.

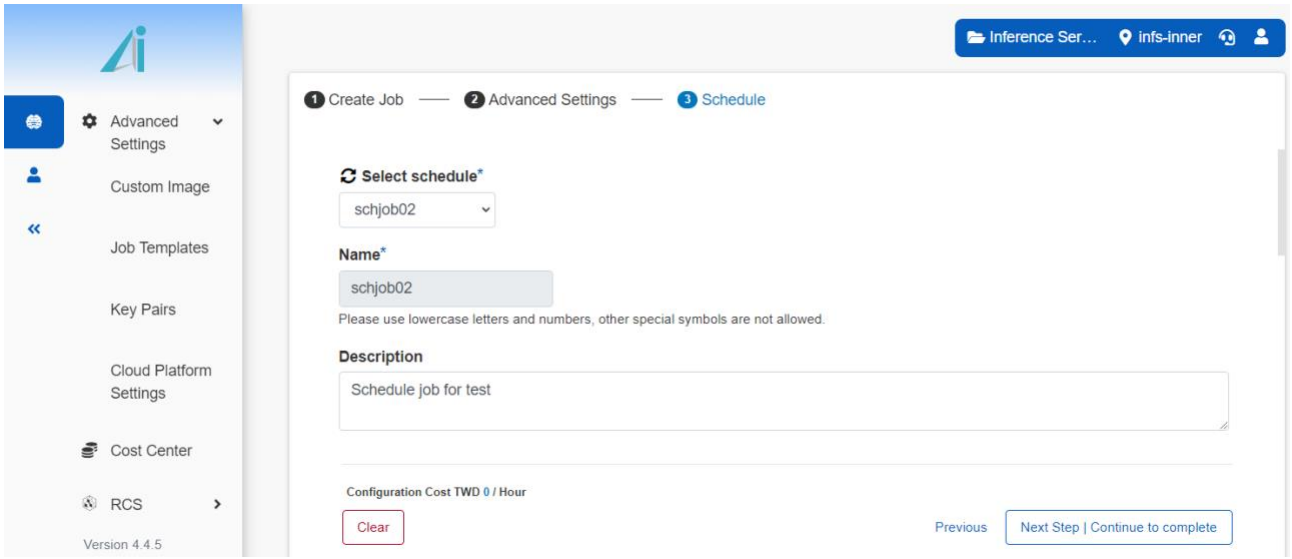
a. Go to "Advanced Settings > Job Template from the sidebar.

b. Select the template and click [Run job with schedule].

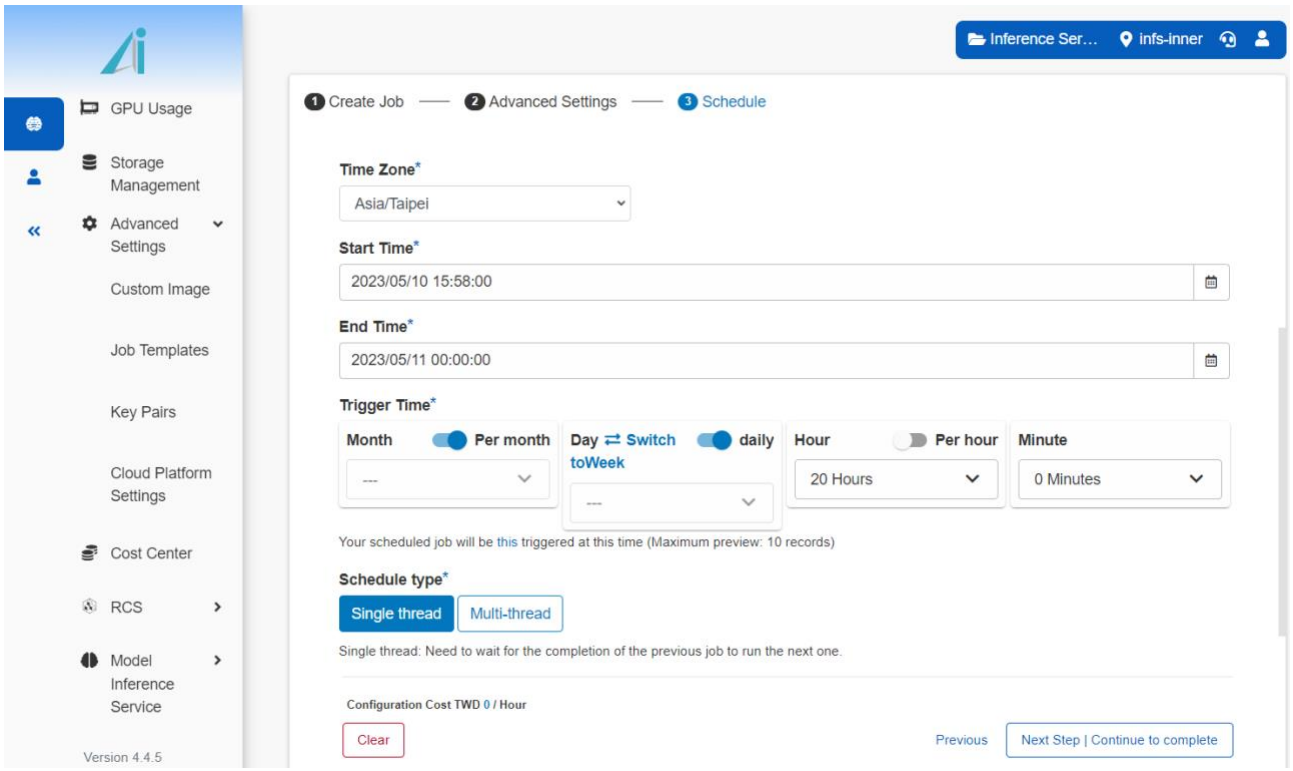


c. Go to job schedule settings, and you can create a schedule, or edit existing schedules.

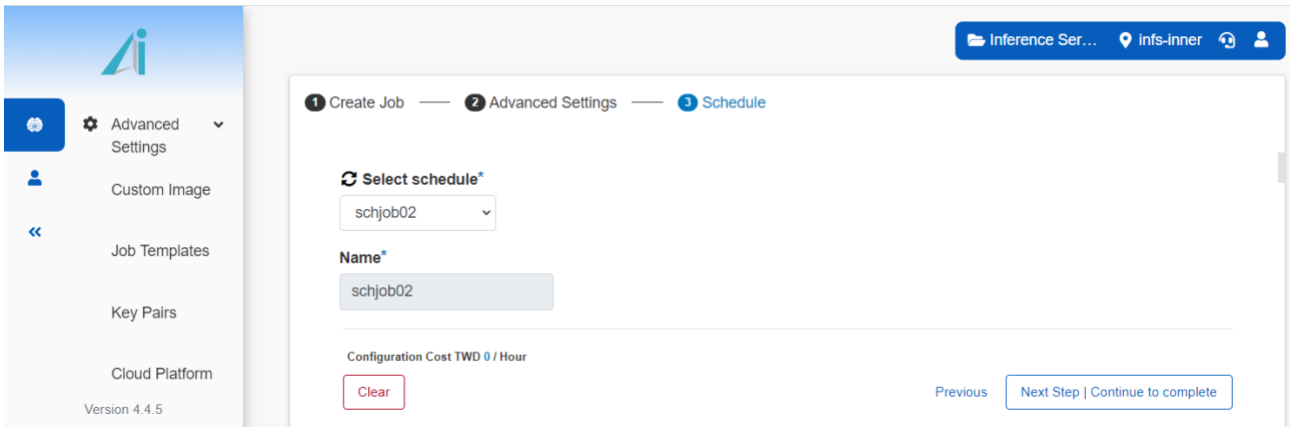
d. If you choose to use existing schedules, the information will be carried out, and you will not be able to change the name of the schedule.


Inference Ser...
infs-inner
Advanced Settings
Custom Image
Job Templates
Key Pairs
Cloud Platform Settings
Cost Center
RCS
Version 4.4.5
1 Create Job
2 Advanced Settings
3 Schedule
Select schedule*
schjob02
Name*
schjob02
Please use lowercase letters and numbers, other special symbols are not allowed.
Description
Schedule job for test
Configuration Cost TWD 0 / Hour
Clear
Previous
Next Step | Continue to complete

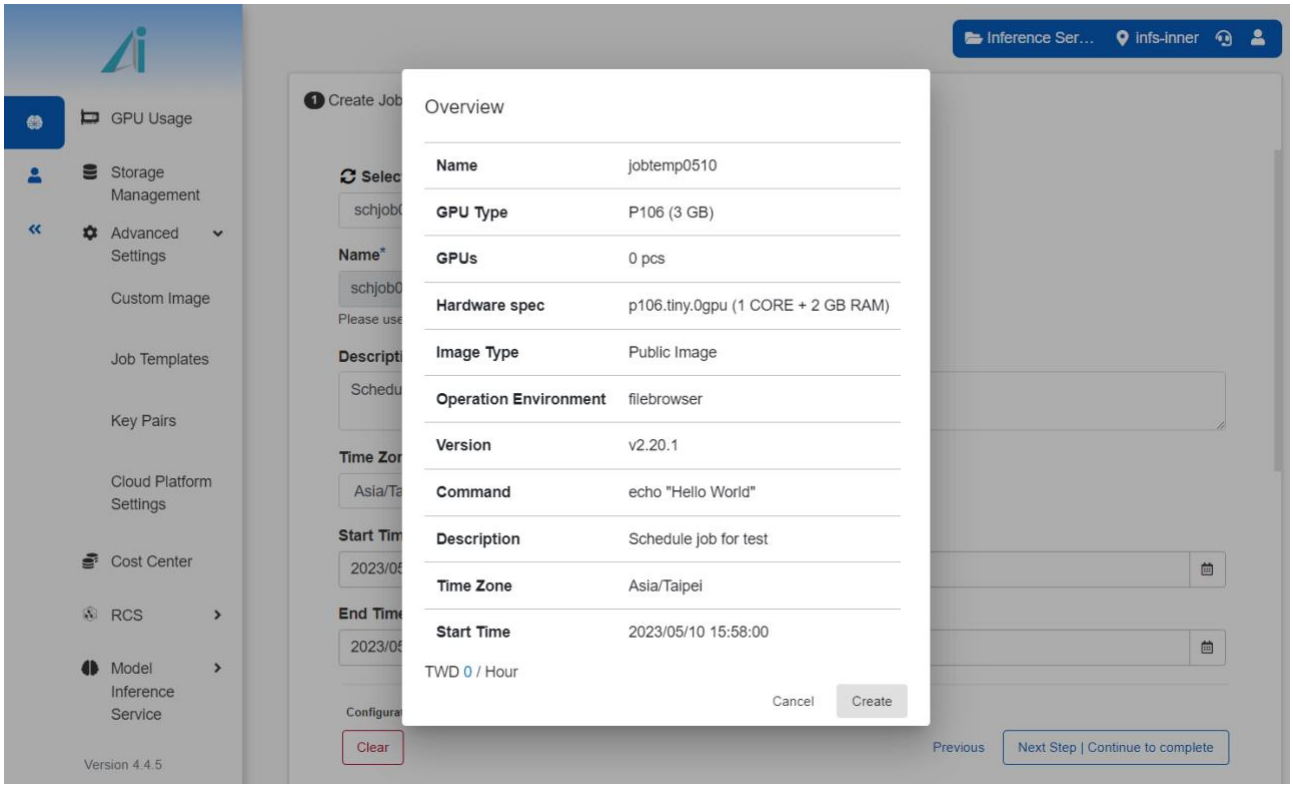
e. After confirming the time zone, start date, end date, initiation time, and schedul e type, click [Next Step | Continue to complete].



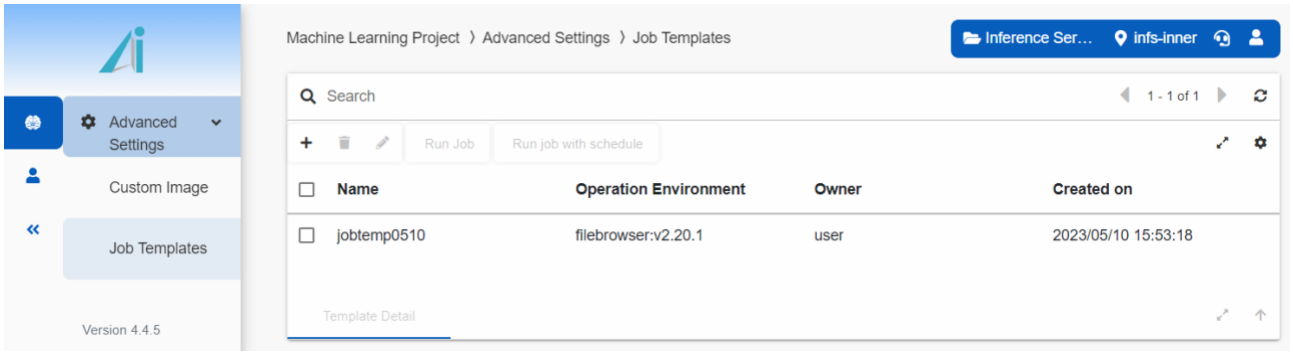
f. If you want to set up a new schedule, you will have to create another one from t he start entering the name, start time, end time, initiation time and schedule typ e. Click [Next Step |Continue to Complete] to proceed.



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



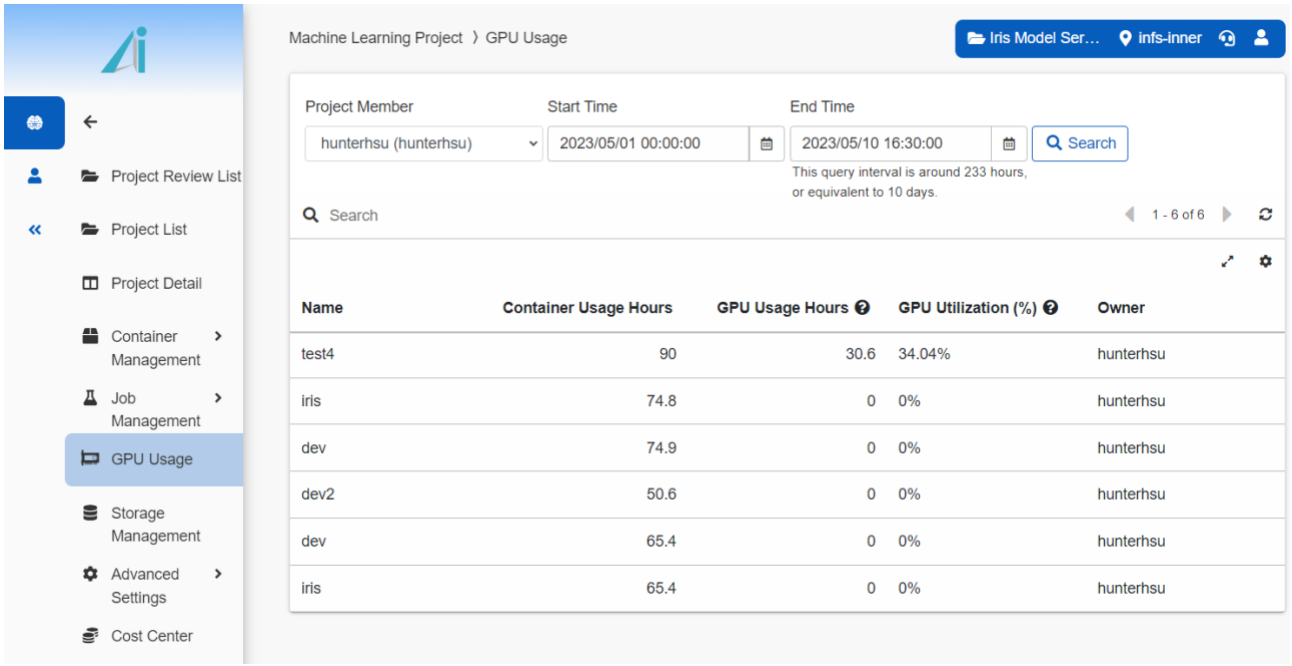
h. The job will be shown on the job list once it is successfully created.



6. GPU Utilization Rate

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

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



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:

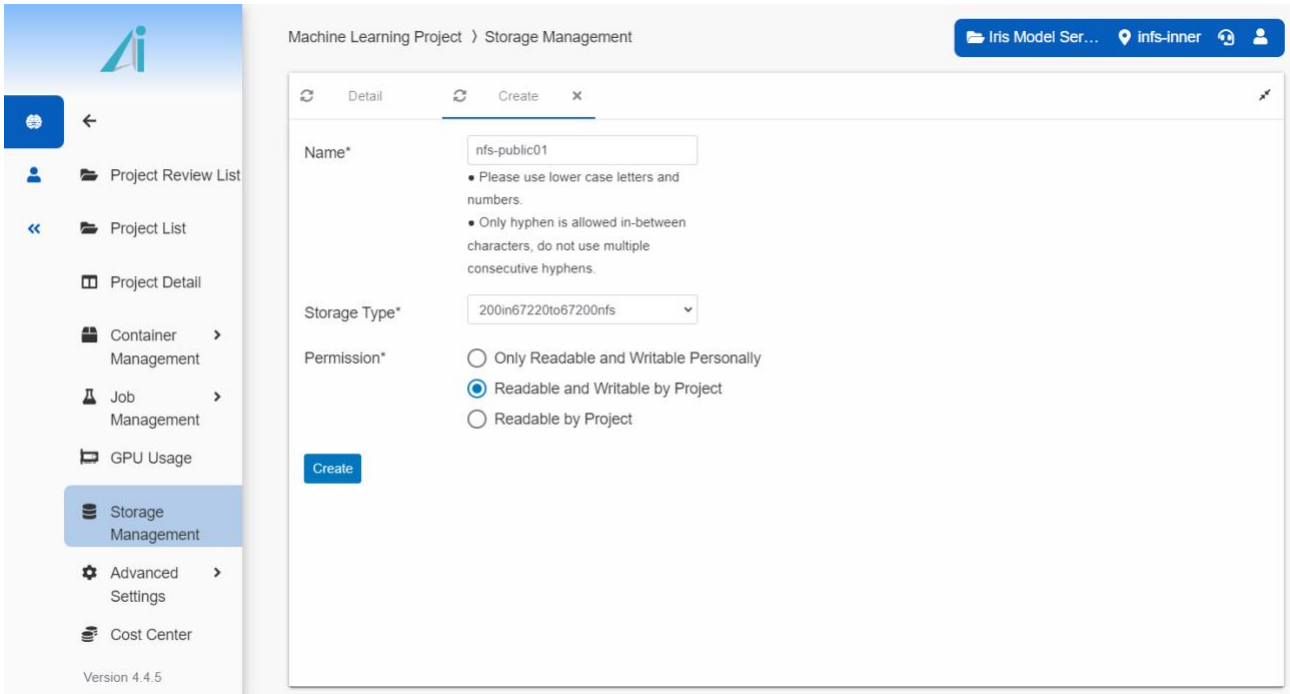
(1) Go to Storage Management from the sidebar.

(2) Click [+]

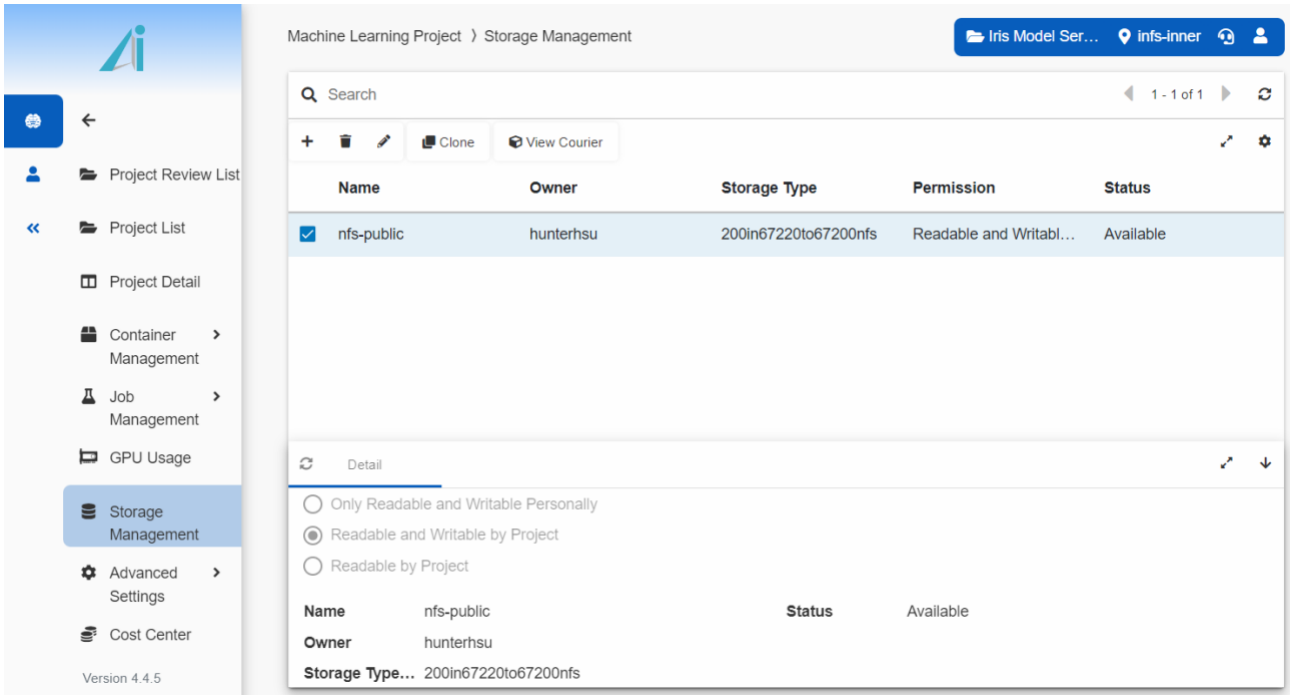
(3) Enter the name and the storage type (storage type is set up by the platform administrator from the admin portal).

(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.



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.

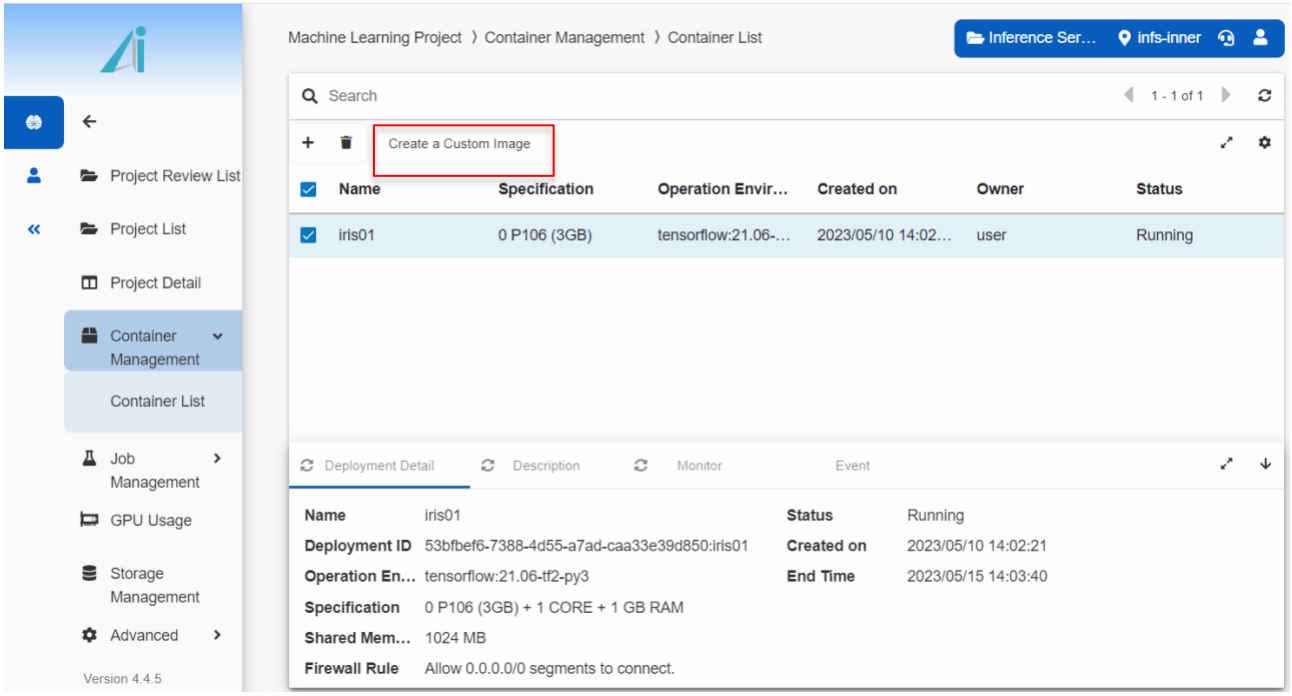
To address this situation, our platform provides a custom image feature. It allows users to preserve the modified container as a template for future use. These custom images can be set for personal use or shared among project members.

Create Custom Image

(1) Go to "Advanced Setting > Custom Image" from the sidebar.

(2) Clink the [+] icon on the top left.

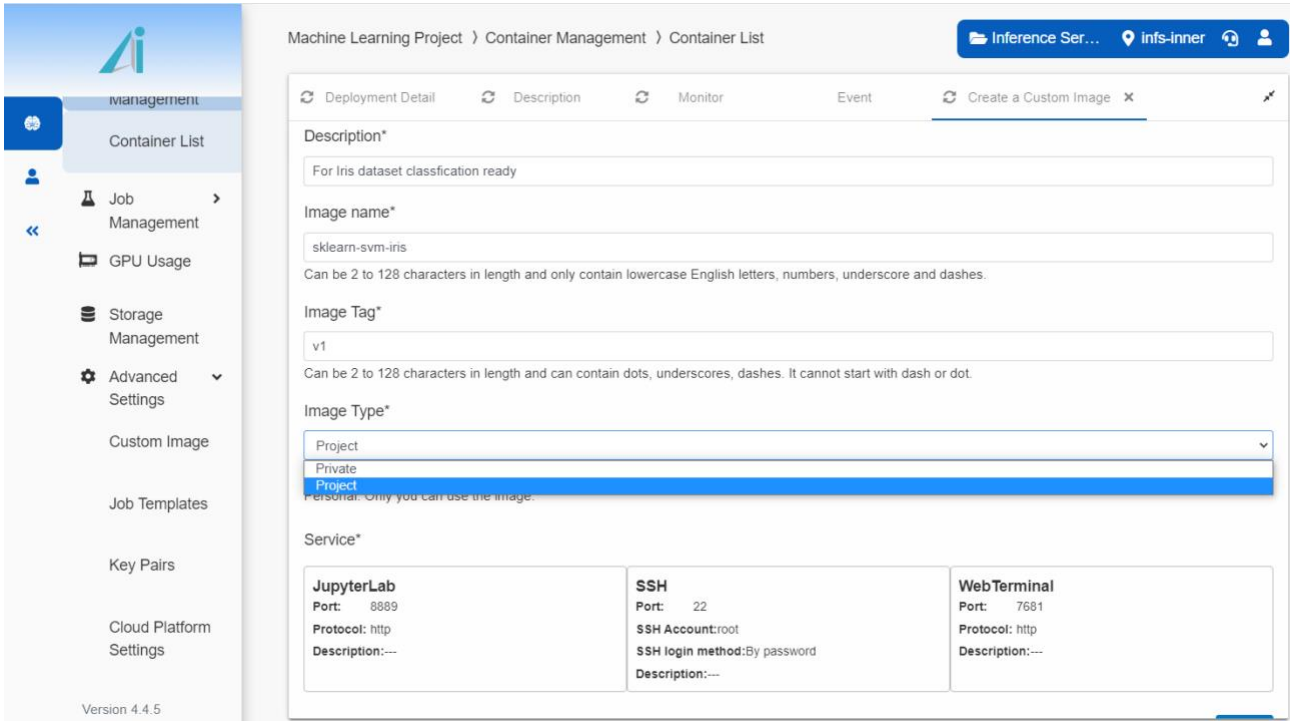
(3) Select the container you wish to set as custom image, and click [Create a Custom Image]


Machine Learning Project 〉 Container Management 〉 Container List
Inference Ser...
infs-inner
Search
1 - 1 of 1
Project Review List
Project List
Project Detail
Container Management
Container List
Job Management
GPU Usage
Storage Management
Advanced
Version 4.4.5
Create a Custom Image
Name
Specification
Operation Envir...
Created on
Owner
Status
iris01
0 P106 (3GB)
tensorflow:21.06-...
2023/05/10 14:02...
user
Running
Deployment Detail
Description
Monitor
Event
Name iris01
Deployment ID 53bfbef6-7388-4d55-a7ad-caa33e39d850:iris01
Operation En... tensorflow:21.06-tf2-py3
Specification 0 P106 (3GB) + 1 CORE + 1 GB RAM
Shared Mem... 1024 MB
Firewall Rule Allow 0.0.0.0/0 segments to connect.
Status Running
Created on 2023/05/10 14:02:21
End Time 2023/05/15 14:03:40

(4) Enter the name and description.

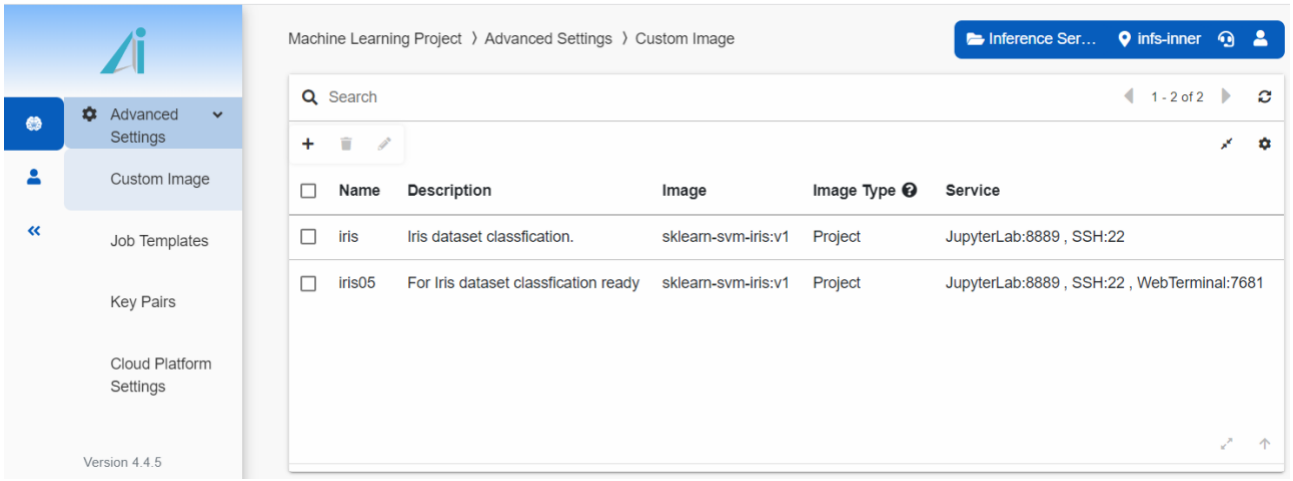
(5) Enter the image name and image tag. (These will sync to the image repository)

(6) Select the image type to set the image for personal use or share with the project.



(7) Confirm the information and click Create.

(8) New image will be displayed in the custom image list.

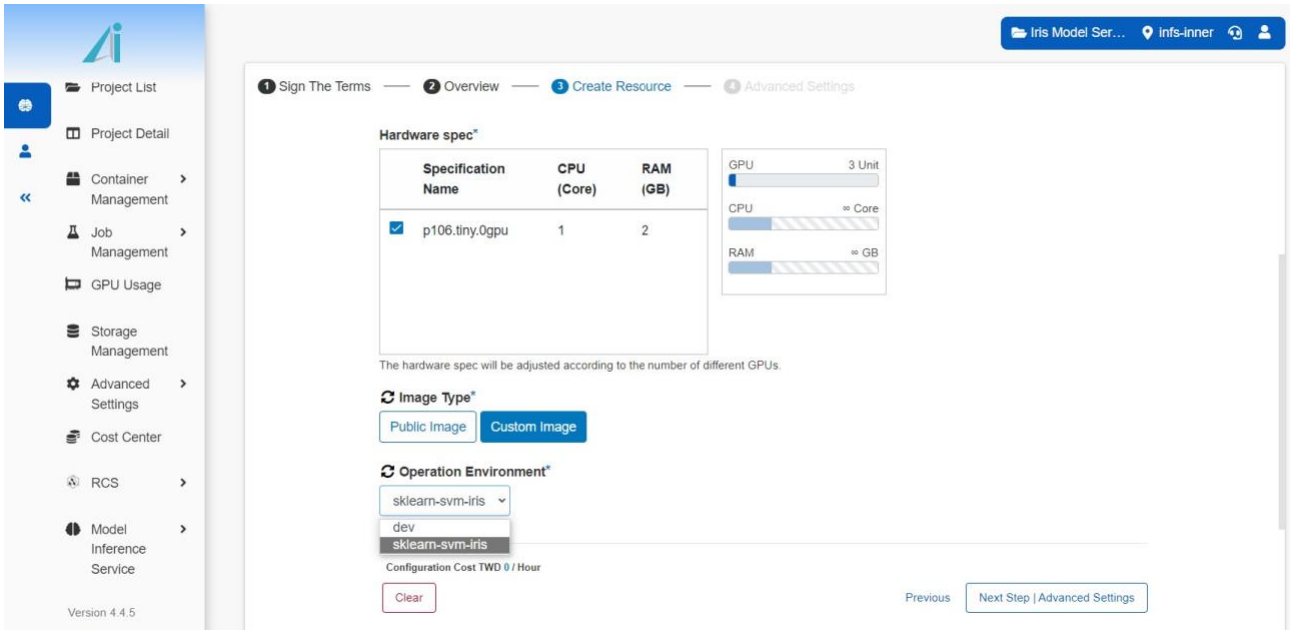


Use Custom Image

(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 fo r you to choose from.



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

8.2 Job Template

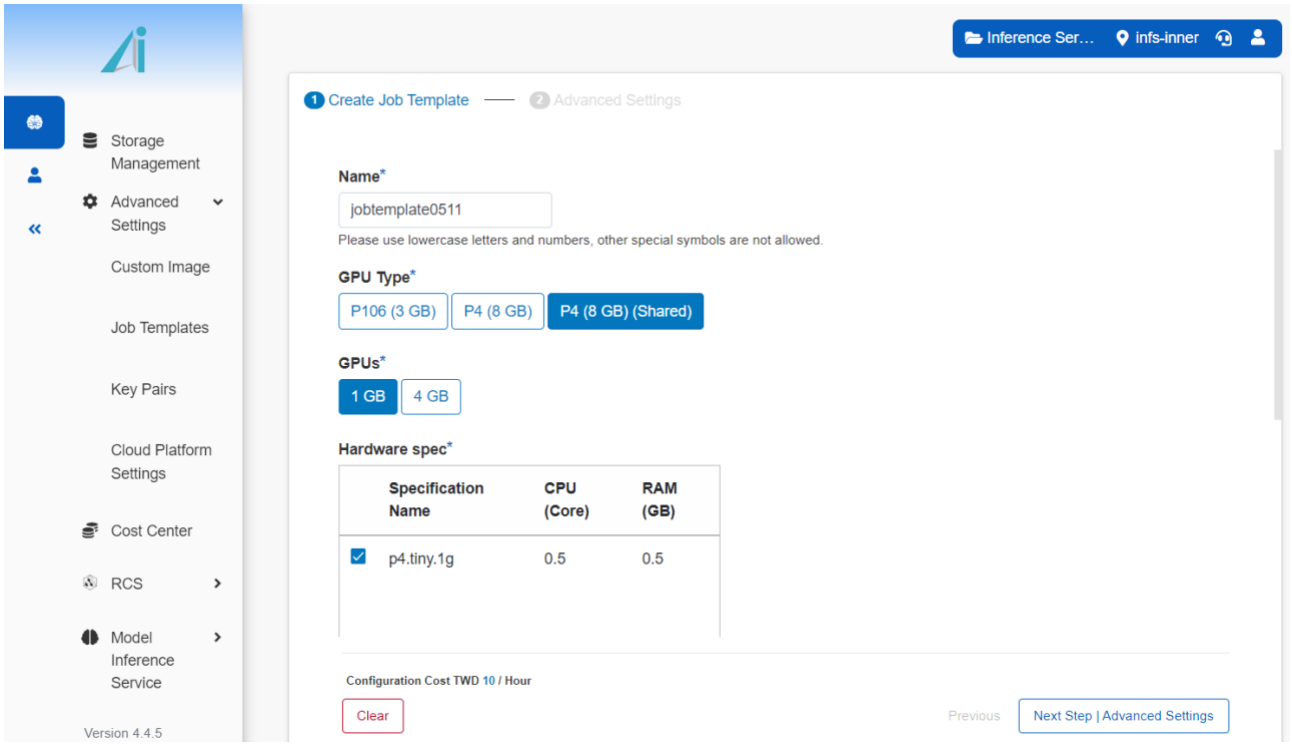
One of the ways to schedule tasks is by pre-defining job templates from scheduling s ettings (*refer to Section 5.4.2 Job Scheduling for more details). By utilizing these pre-defi ned job templates, jobs can be directly launched and executed.

Create Job Template

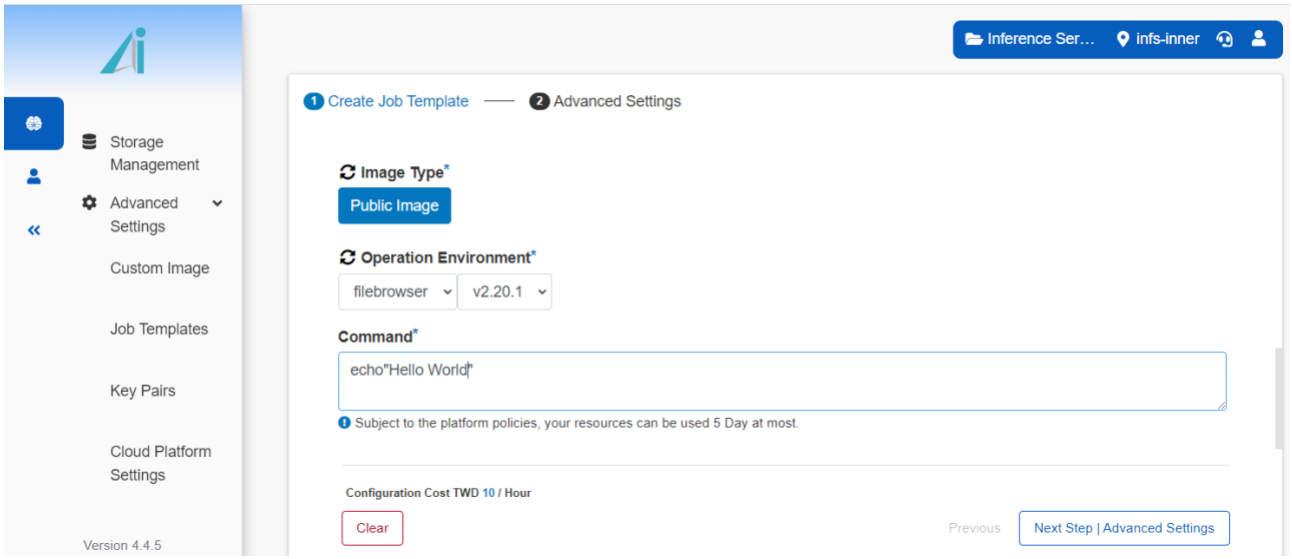
(1) Go to "Advanced Setting > Job templates" from the sidebar

(2) Click [+] on the top left corner.

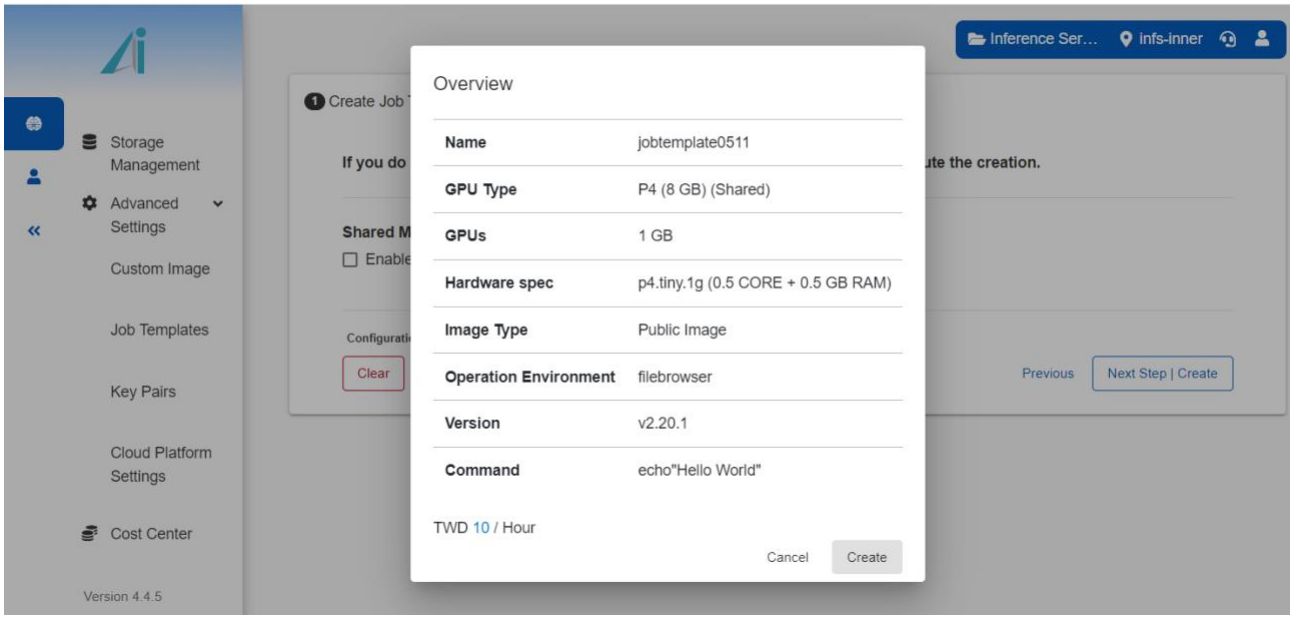
(3) Enter the name, GPU type, GPU pcs, Hardware Spec.



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

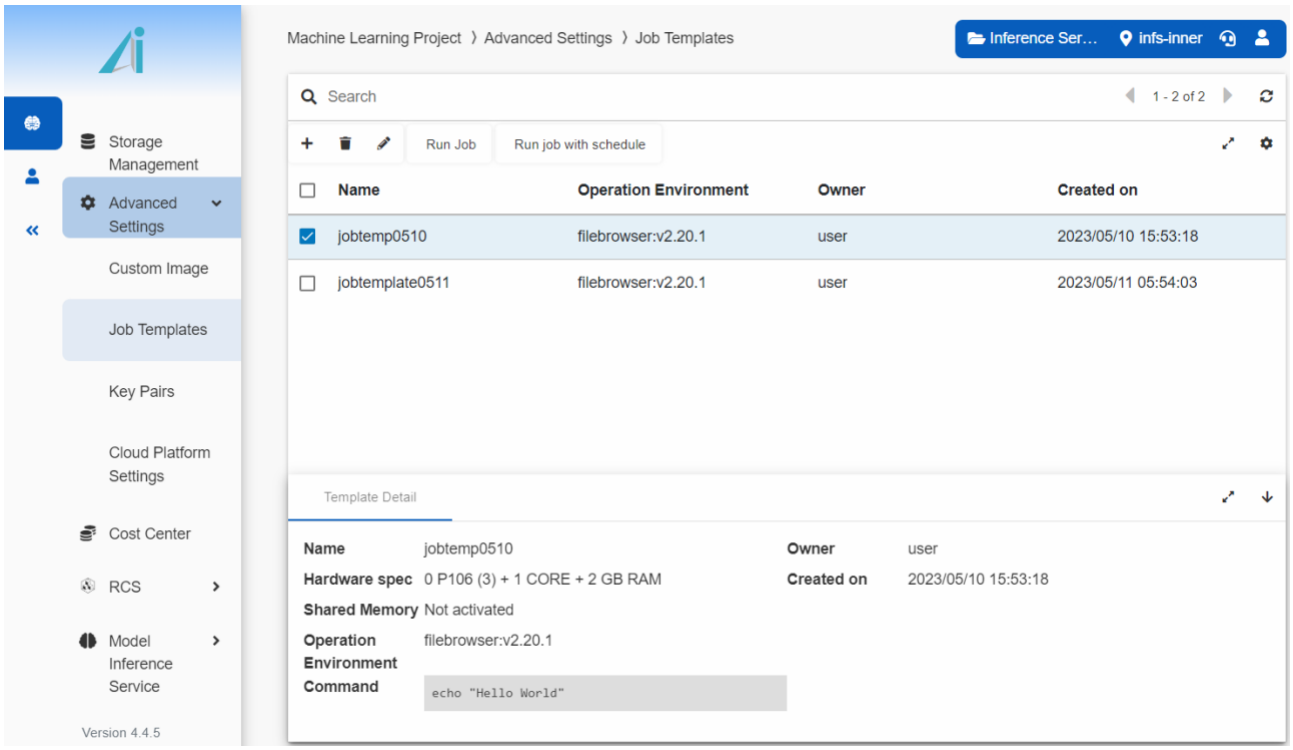


(5) Confirm the entered information on the overview page, click [create] to proceed.

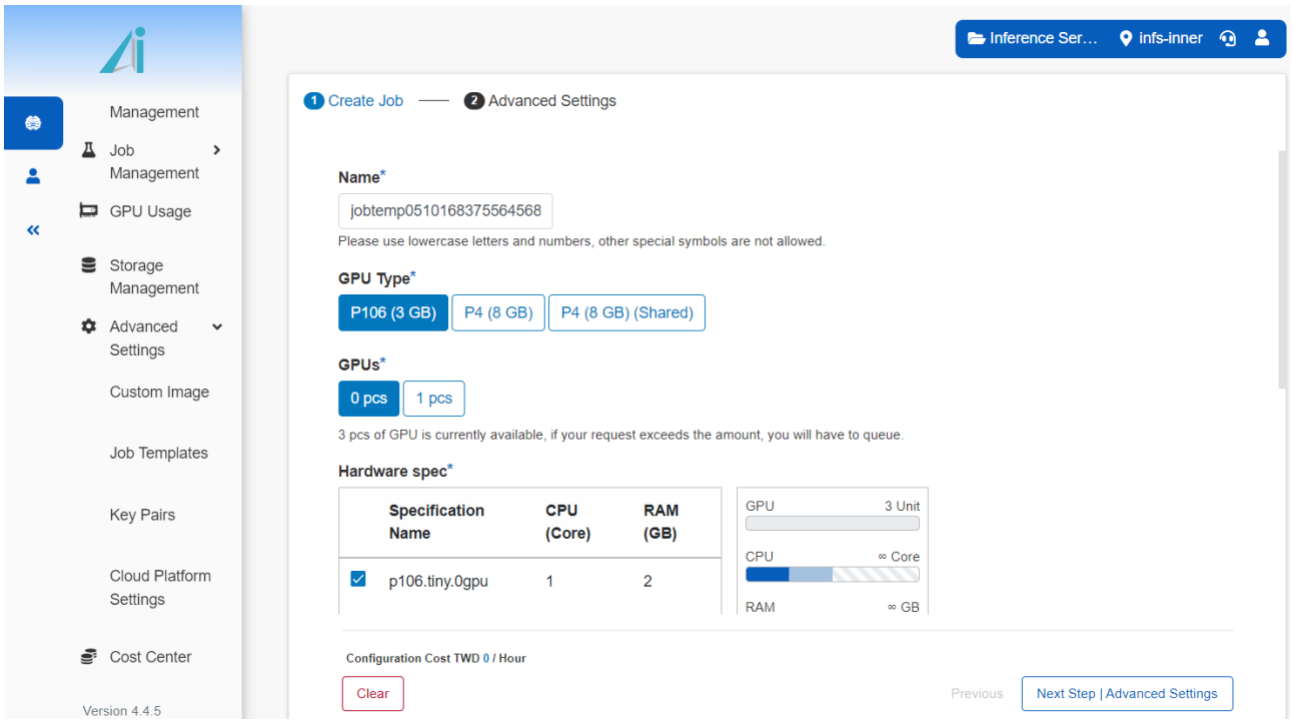


Use Job Template

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

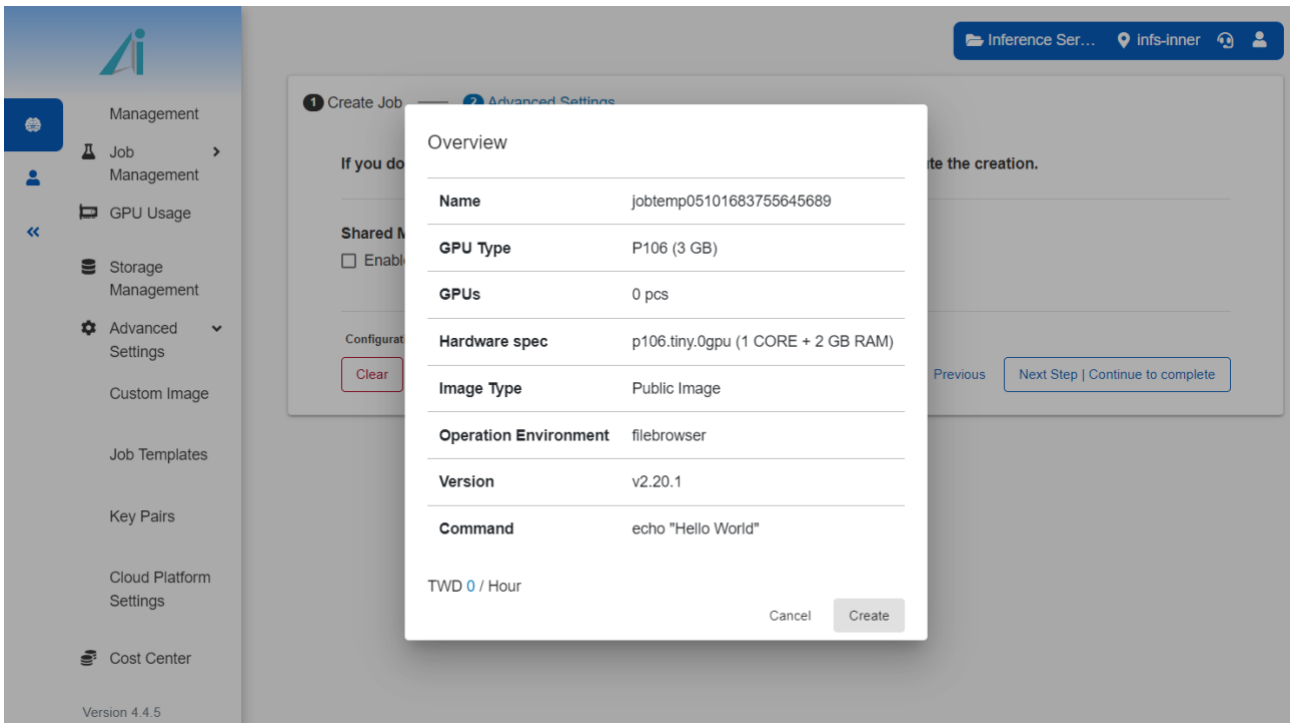


(2) The information is automatically filled into the job creation page, where the job name is default to the template name followed by a serial number. You can edit the name at your will.

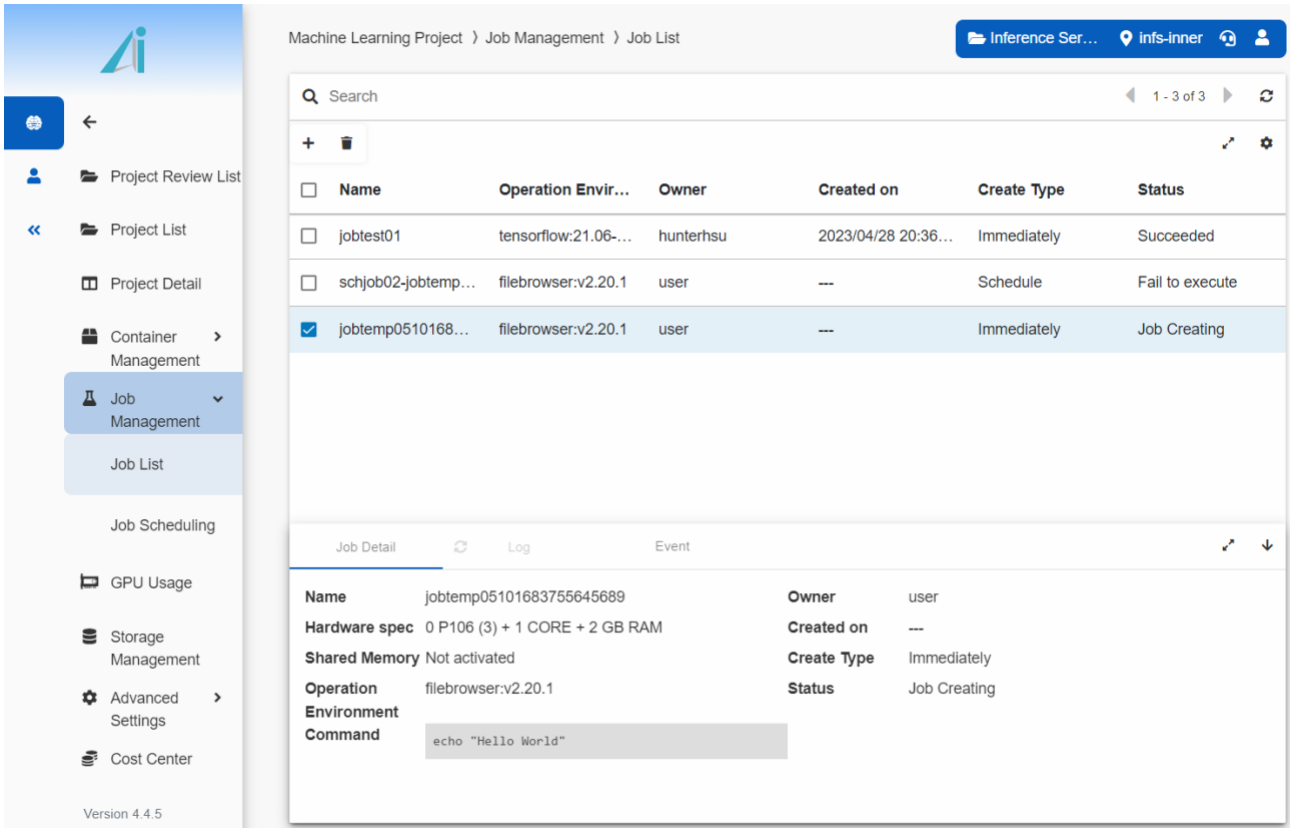


(3) Click [Next | Advanced Setting]

(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…
infs-inner
Search
1 - 3 of 3
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
Name
Operation Envir…
Owner
Created on
Create Type
Status
jobtest01
tensorflow:21.06-…
hunterhsu
2023/04/28 20:36…
Immediately
Succeeded
schjob02-jobtemp…
filebrowser:v2.20.1
user

Schedule
Fail to execute
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

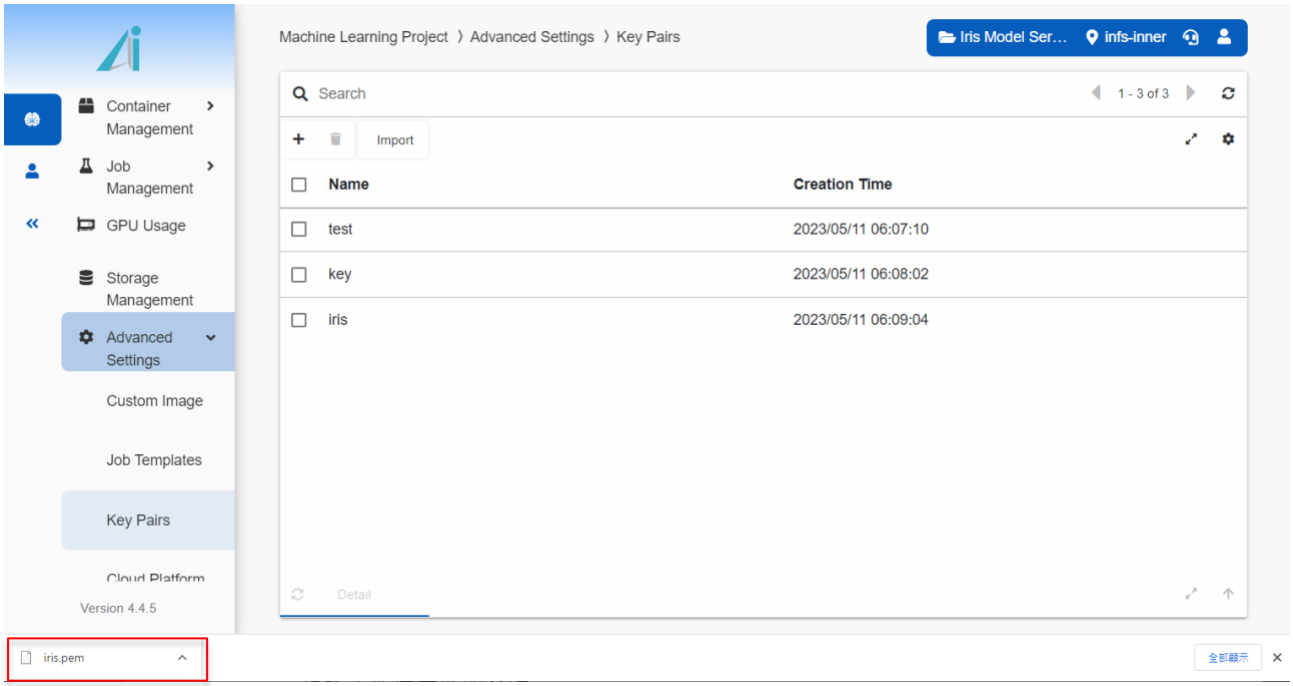
Shared Memory
Not activated
Create Type
Immediately
Operation Environment
filebrowser:v2.20.1
Status
Job Creating
Command
echo "Hello World"

8.3 Key Pair

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

Create Key Pair

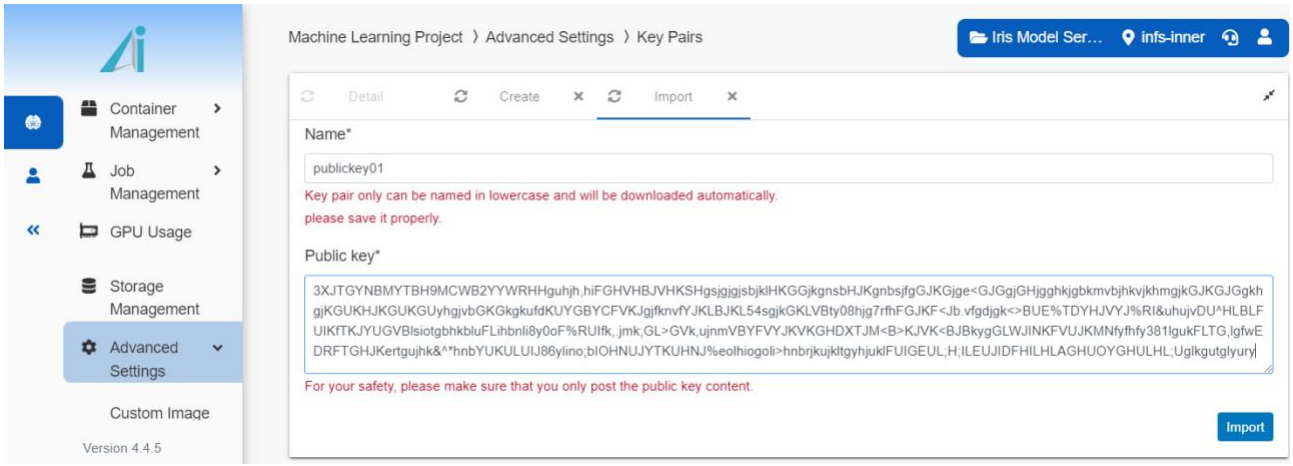
The key pair created by the users are for SSH connection for MLS services. Go to “Advanced Settings > Key Pair” and create a new key pair by clicking on the [+] on the top left and download the private key to the computer of the users.



※ 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.

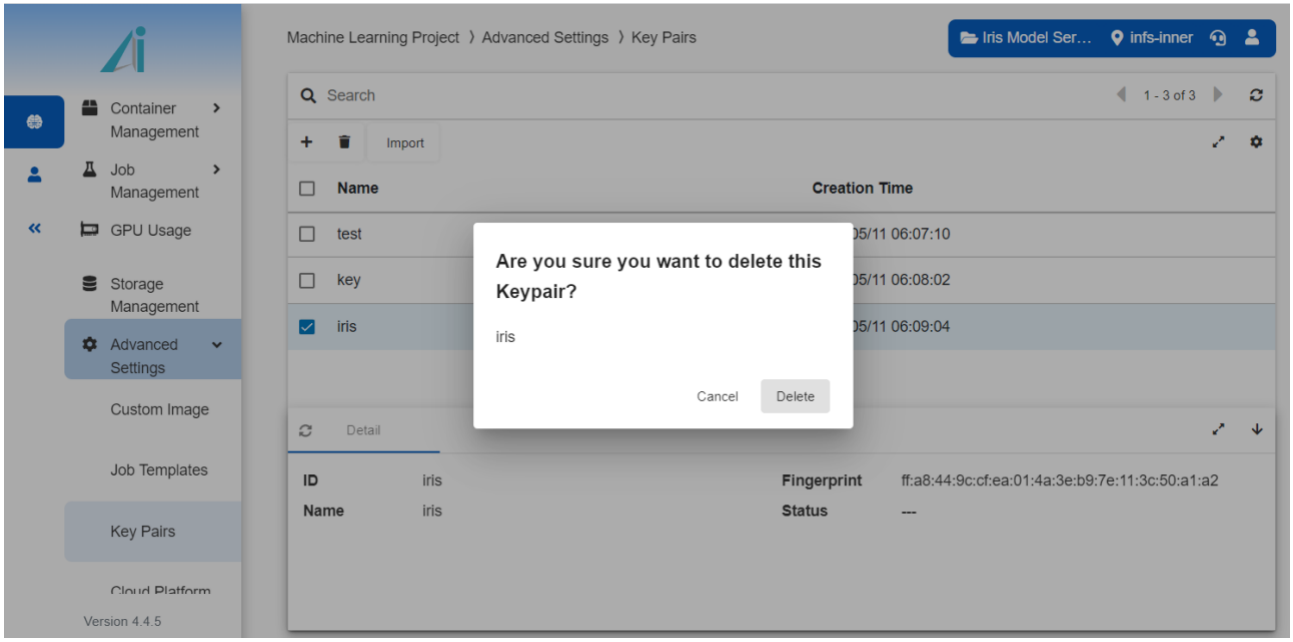
Import Keys

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



Delete Keys

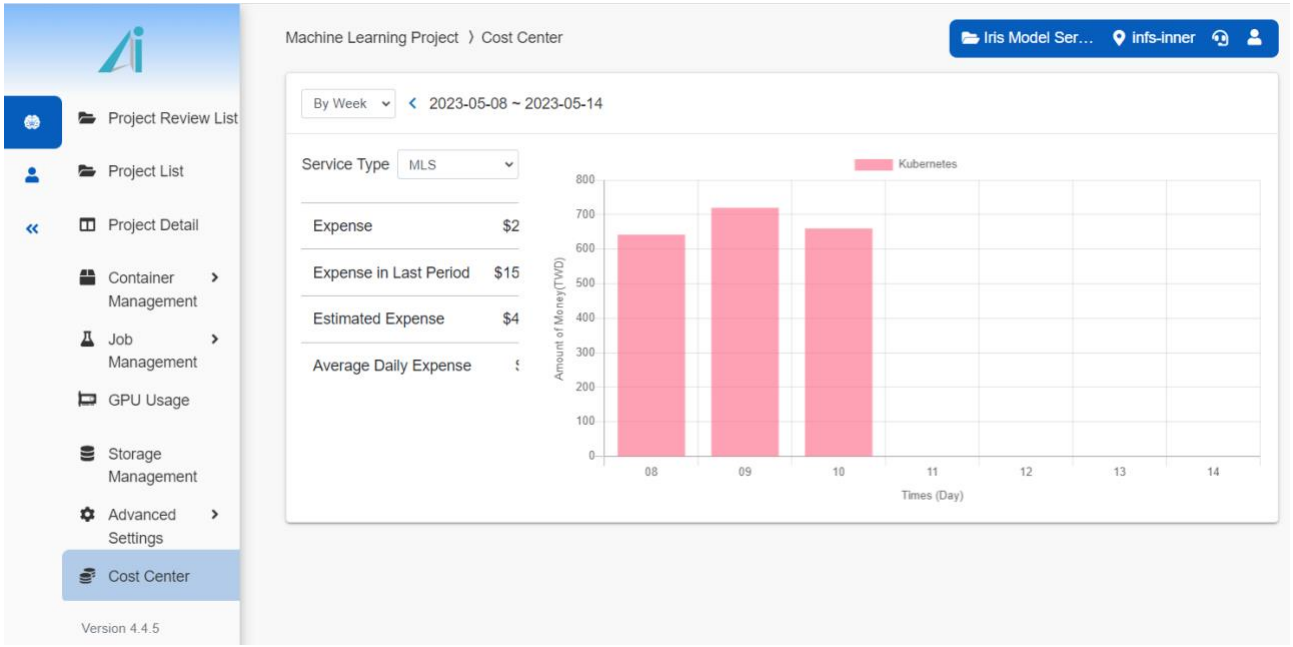
To delete a key, select the key you wish to delete from the list and click on the icon . A confirmation will appear, as shown in the following figure. If everything is correct, click [Dele te] to proceed with the deletion.



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

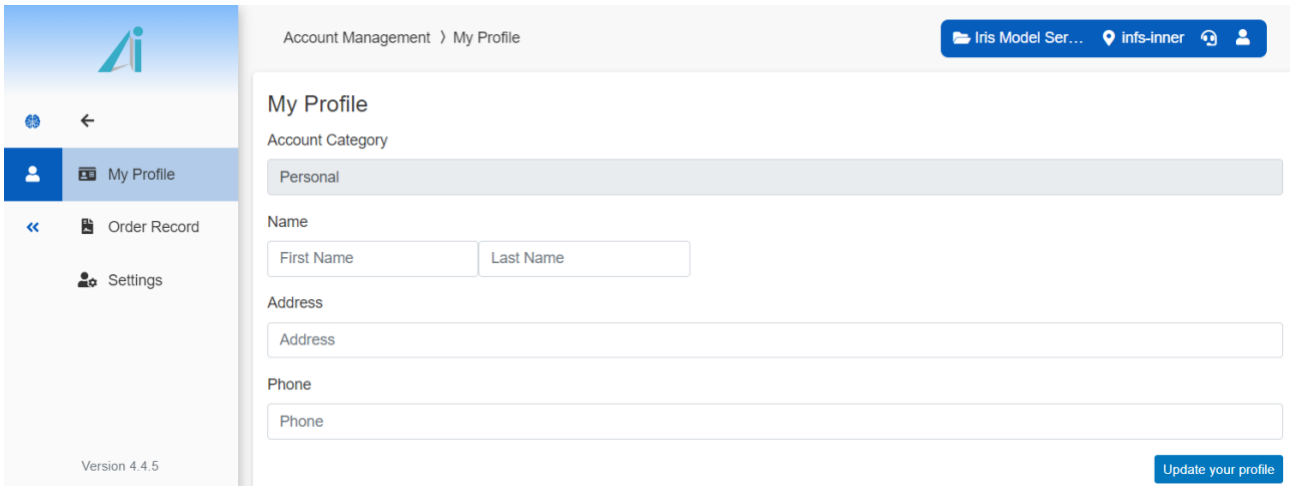
*The hourly cost of each container is set up by the platform administrators from the a dmin protal.



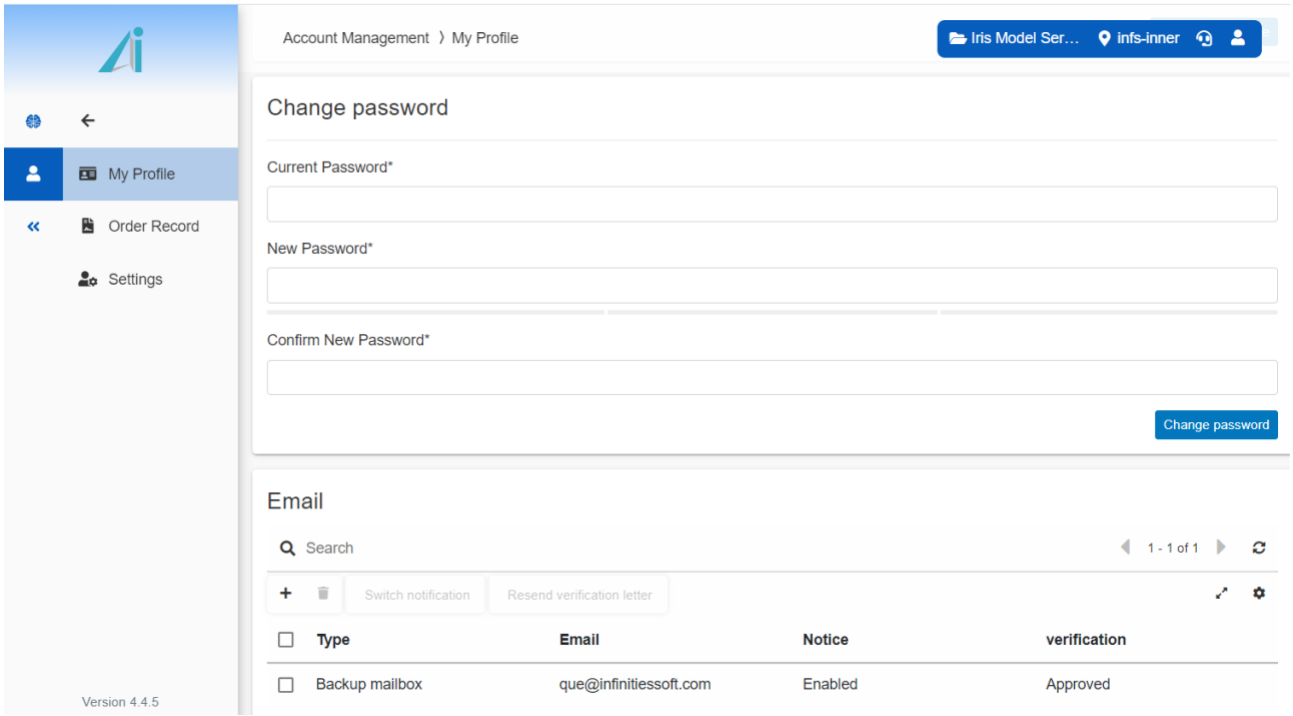
6. Account Management

1. General Information

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

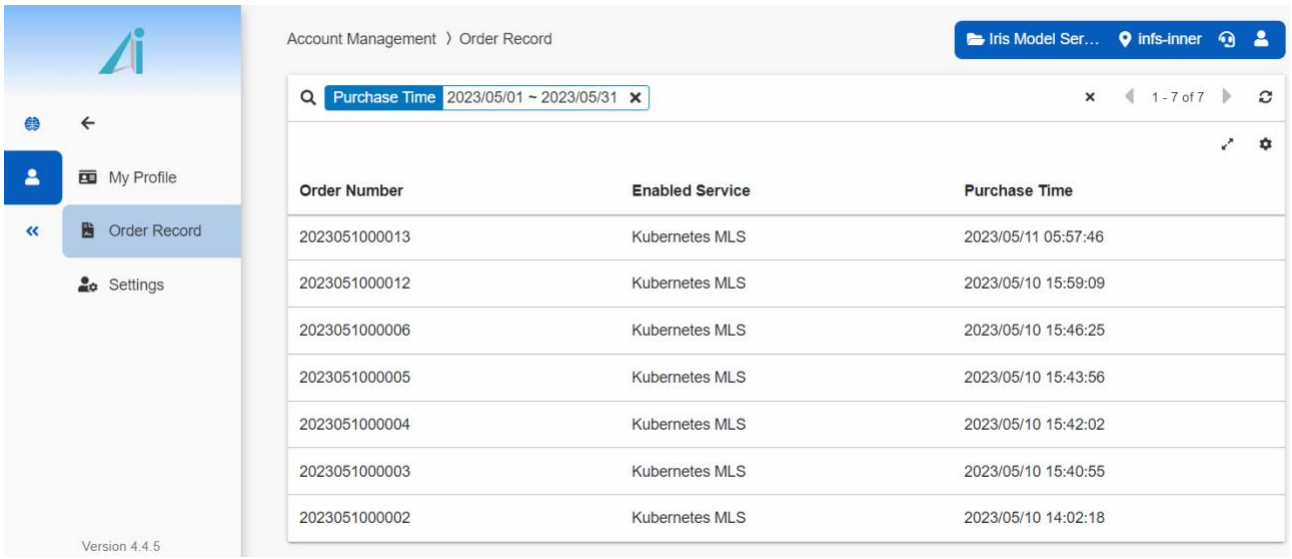


Other features include changing your password and email address, turn on/ off notificati ons and resending confirmation emails.

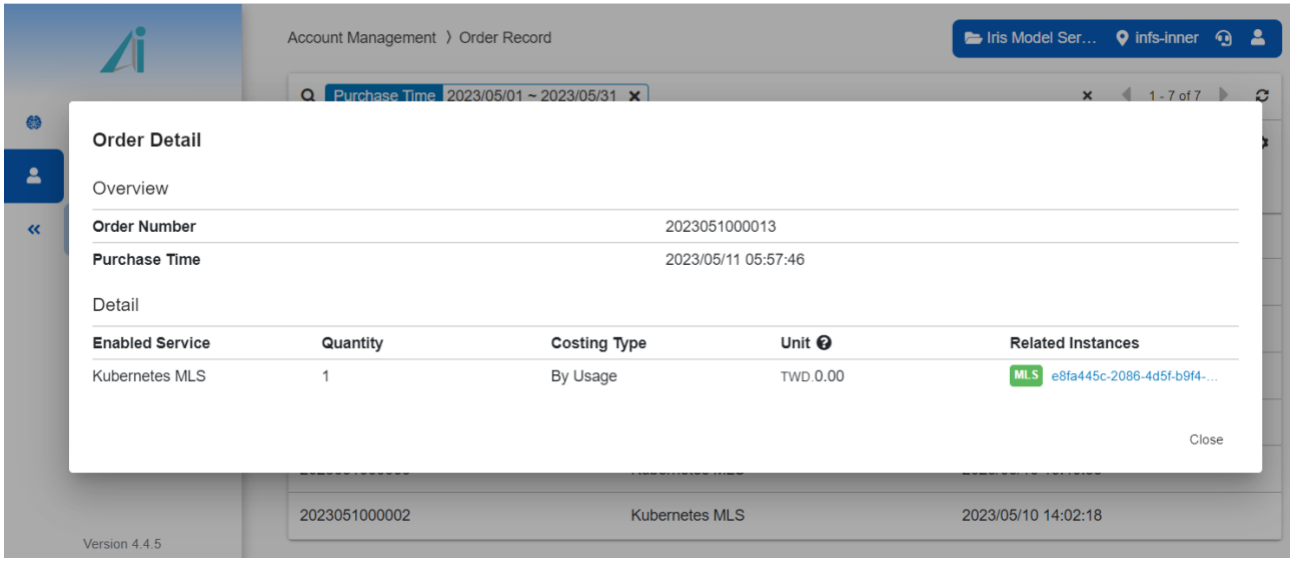


2. Order Record

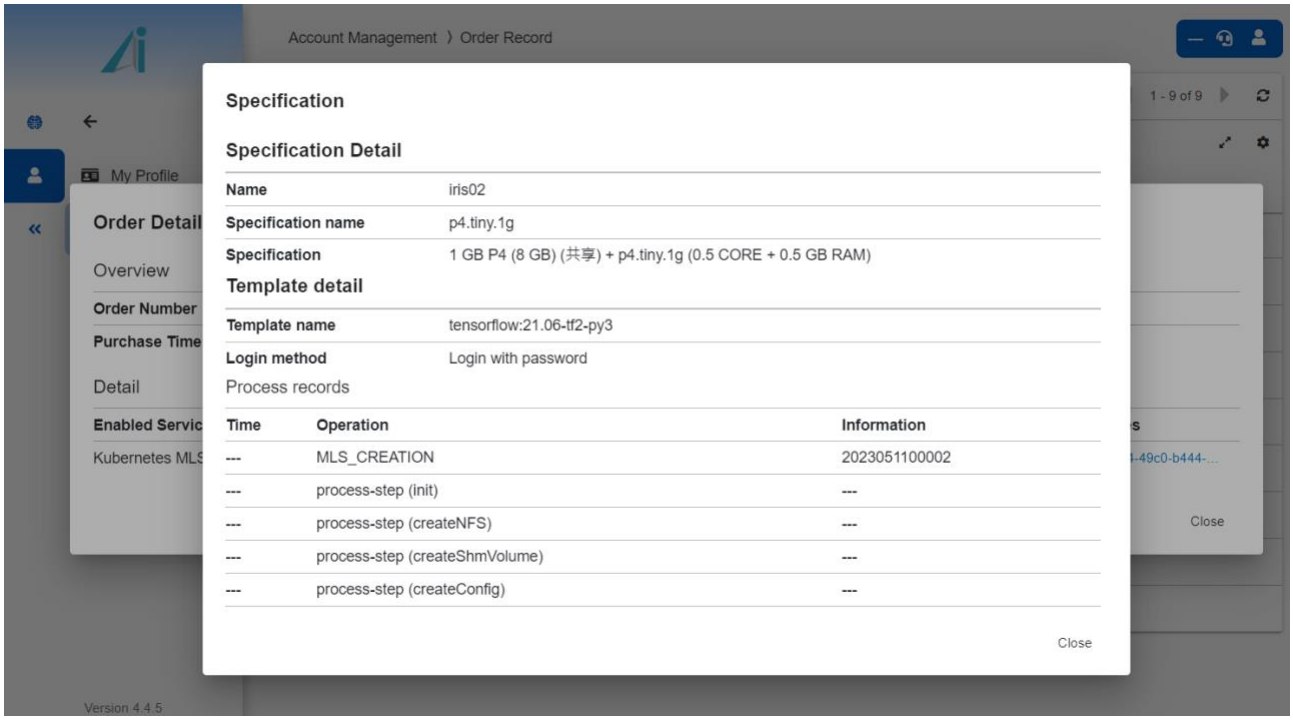
This feature allows each account to view the services created during a period of time.

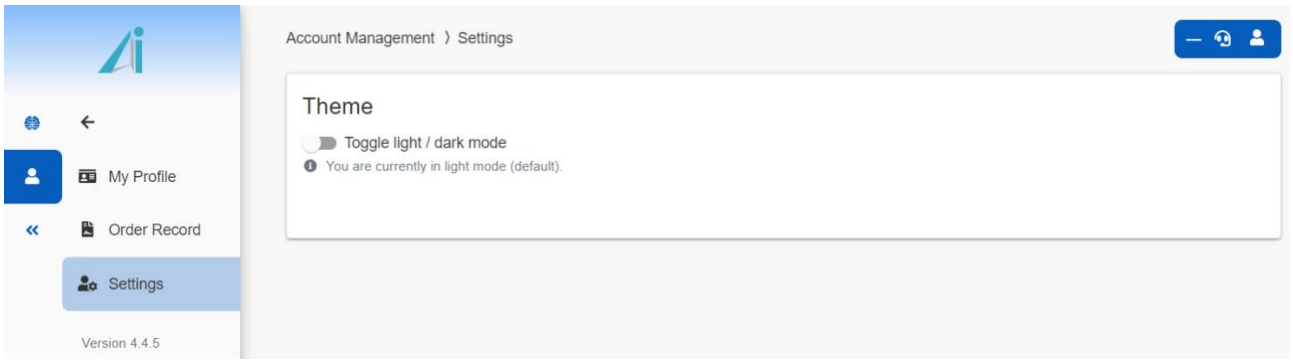


Detail of an order is available through selecting a period of the [Purchase time] and click on a specific order.



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



3. Settings

Switch between Light or dark mode.