



雲林科技大學AI雲平台-系統使用說明

Outline

- 淺談虛擬化技術：虛擬機與容器
- AI雲平台系統示意圖
- AI雲平台名詞解說
- AI雲平台實際操作說明

淺談虛擬化技術：虛擬機與容器

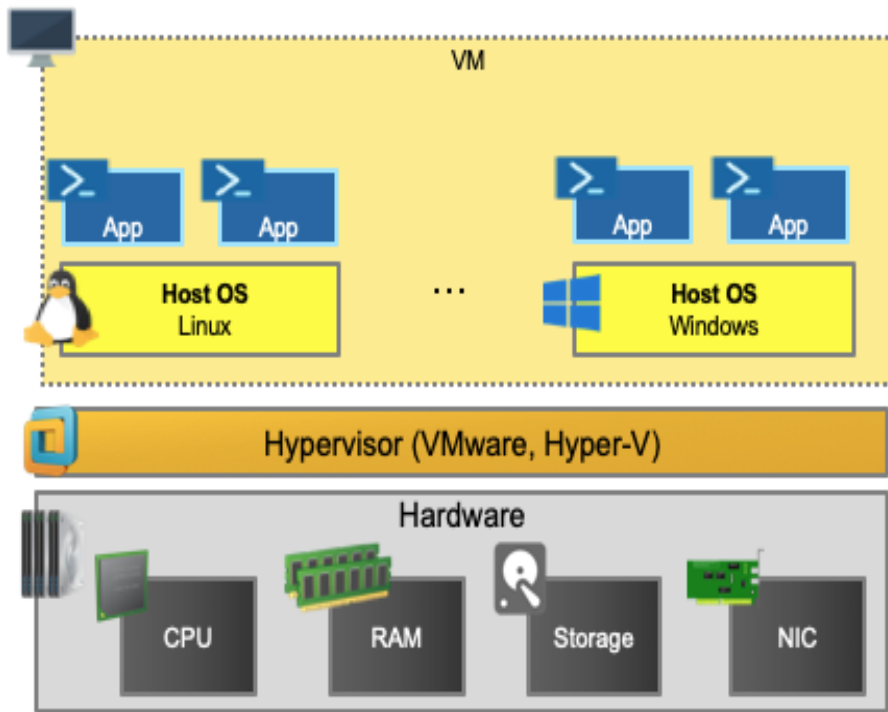
虛擬機(VM)與容器(Container)

作業系統(OS)虛擬化技術，
處理如何分享硬體資源的技術

作業系統虛擬化技術，
通常會有一個Hypervisor的角色，負責管理與
分配硬體資源

雲級別的Hypervisor：AWS、GCP、Azure..

作業系統級別的Hypervisor：VirtualBox、
VMWare、Hyper-V...

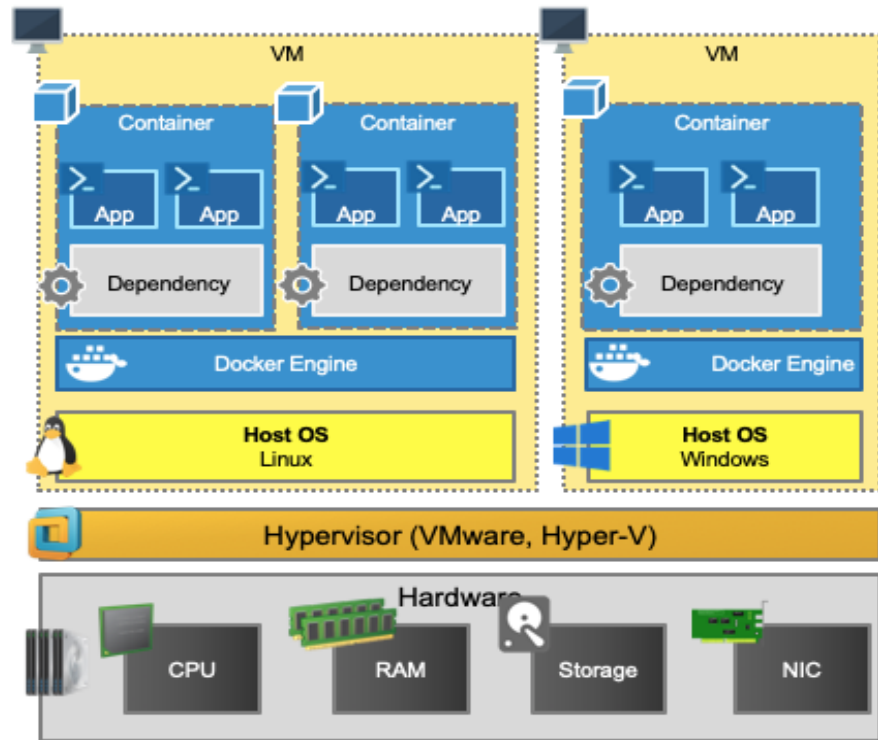


虛擬機(VM)與容器(Container)

容器(Container)虛擬化技術，
處理軟體系統的環境建置

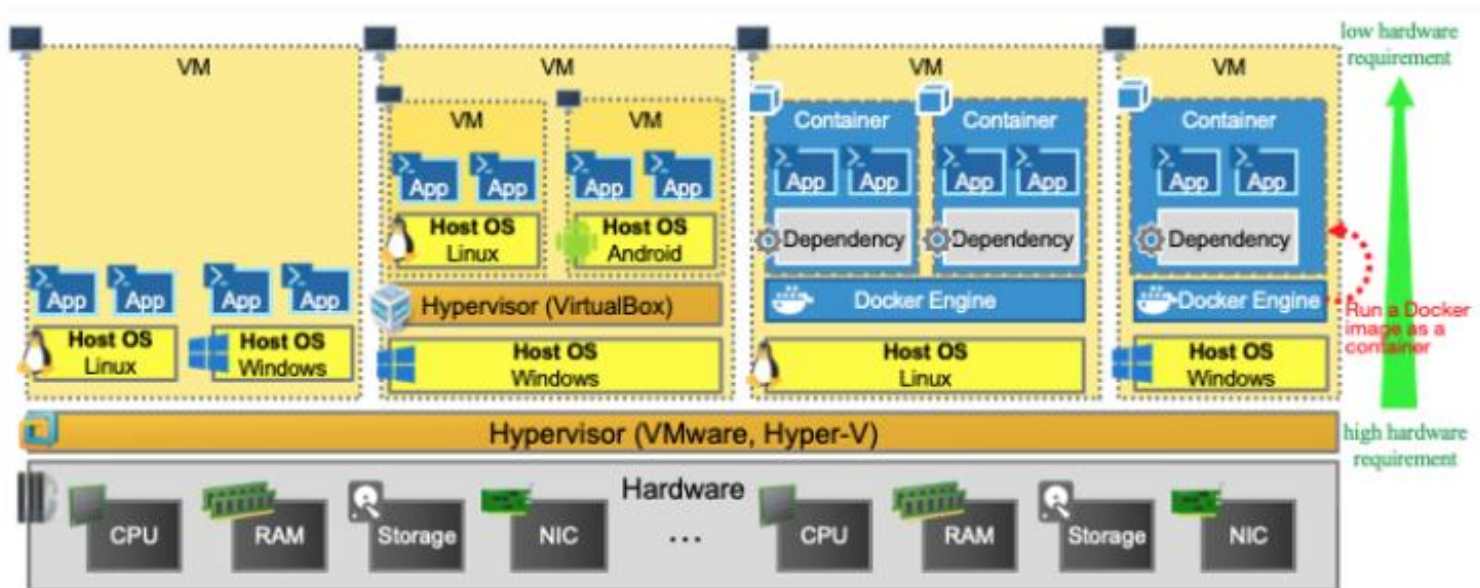
讓整個軟體系統的執行環境可以搬遷，
「與其為每一個軟體系統建置執行環境，倒不
如讓整個軟體系統的執行環境是可以搬遷
與重複利用的」

容器(Container)就是一種可以搬遷與重複利
用的執行環境



虛擬機(VM)與容器(Container)

容器化技術並不是用來取代傳統作業系統虛擬化技術



虛擬機與容器-兩種虛擬化技術的差異

- 以代表性服務為例：**Docker**是以**App**為單位的虛擬化技術；**VMWare**是以作業系統為單位的虛擬化技術。
- 容器虛擬化技術不是用來取代作業系統虛擬化技術的新技術
- 容器並不是**VM**
- 容器虛擬化技術以共享**Host OS Kernel**的方式來執行**App**
- 容器(**container**)是執行環境的實例(**instance**)；**VM**是作業系統的實例(**instance**)
- 容器化技術更面向**DevOps**的需求
- 硬體資源的分配
- 應用程式的獨立性

AI雲平台-系統示意圖

AI雲平台-系統示意圖



Admin

透過 Gemini AI Console
Web管理平台資源並可設
定資源排程

AI雲平台



容器

Site(常駐型容器)、Job(任務型容器)、JupyterToJob(動態取用GPU型容器)



User

1. 使用者登入AI雲平台Web即可隨時取得的容器資源。
2. 可從Web展開容器提供的IDE介面的新瀏覽器視窗，進行課程學習。如Jupyternotebook



NAS

3. 連接與容器相同的資料空間，進行檔案存取

H100 GPU
Server



A100 GPU
Server



GPU Server
(可擴充)



GPU Server
(可擴充)

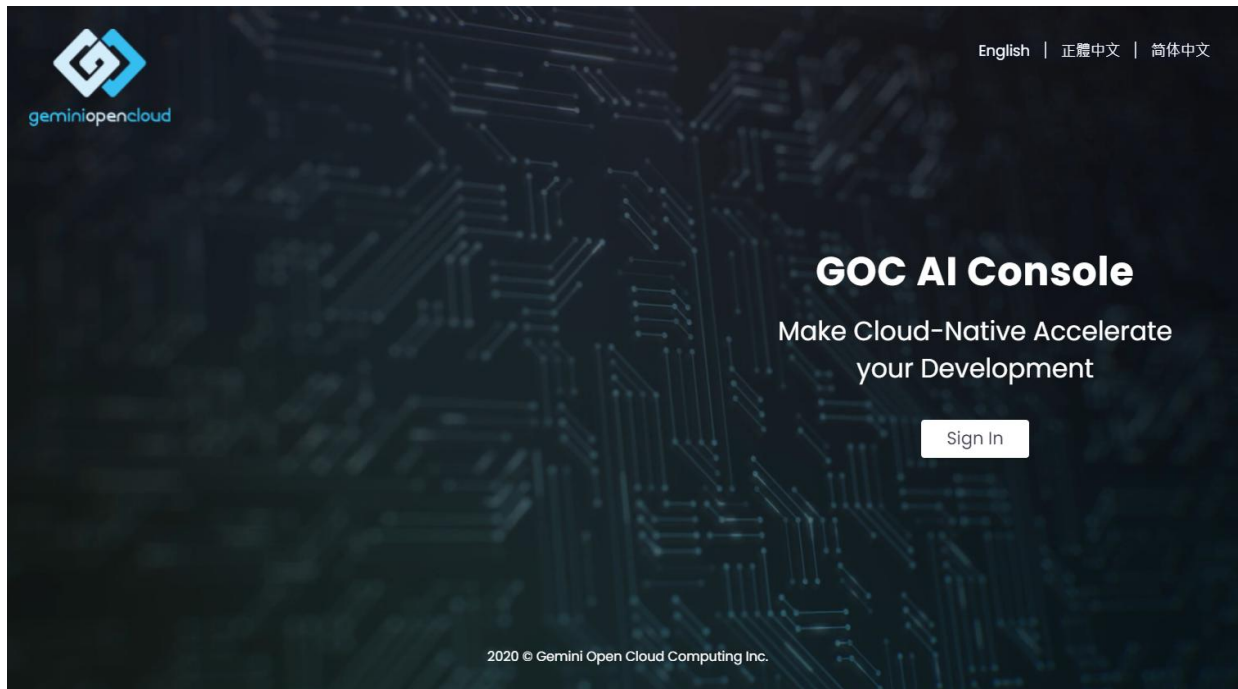


AI雲平台名詞解說

AI雲平台-Portal

提供管理員資源管理

提供使用者容器資源取得與使用



AI雲平台-Application(常駐型容器)

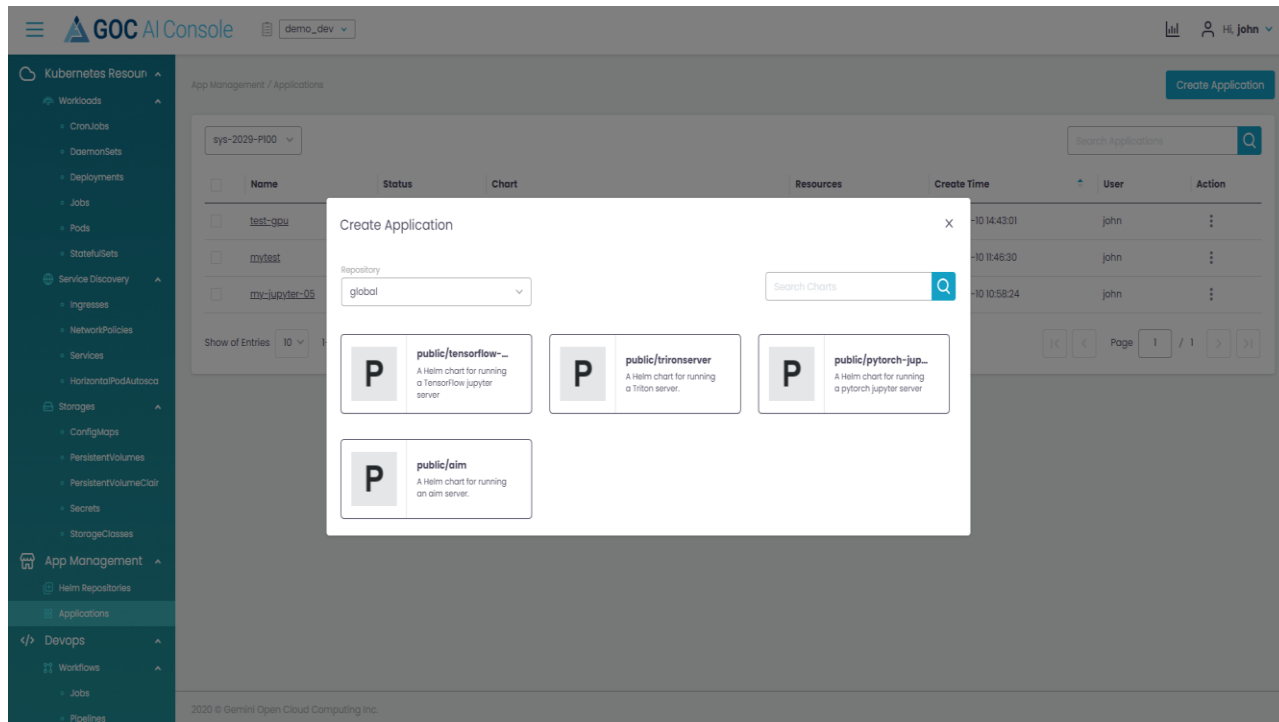
可依需求自行建立容器

依容器服務特性提供連線資訊

提供網頁式的編程開發工具介面，方便使用者快速取得工具

網頁式的編程開發工具，方便程式開發人員即時運程式，即時除錯

可將程式碼儲存至個人空間，利用Job取得更多運算資源執行程式



AI雲平台-JupyterToJob(動態取用GPU型容器)

以常駐型容器為基礎

建立時無須占用GPU資源

程式執行時，可選擇執行環境
資源規模，動態取用GPU

AI雲平台-Job(任務型容器)

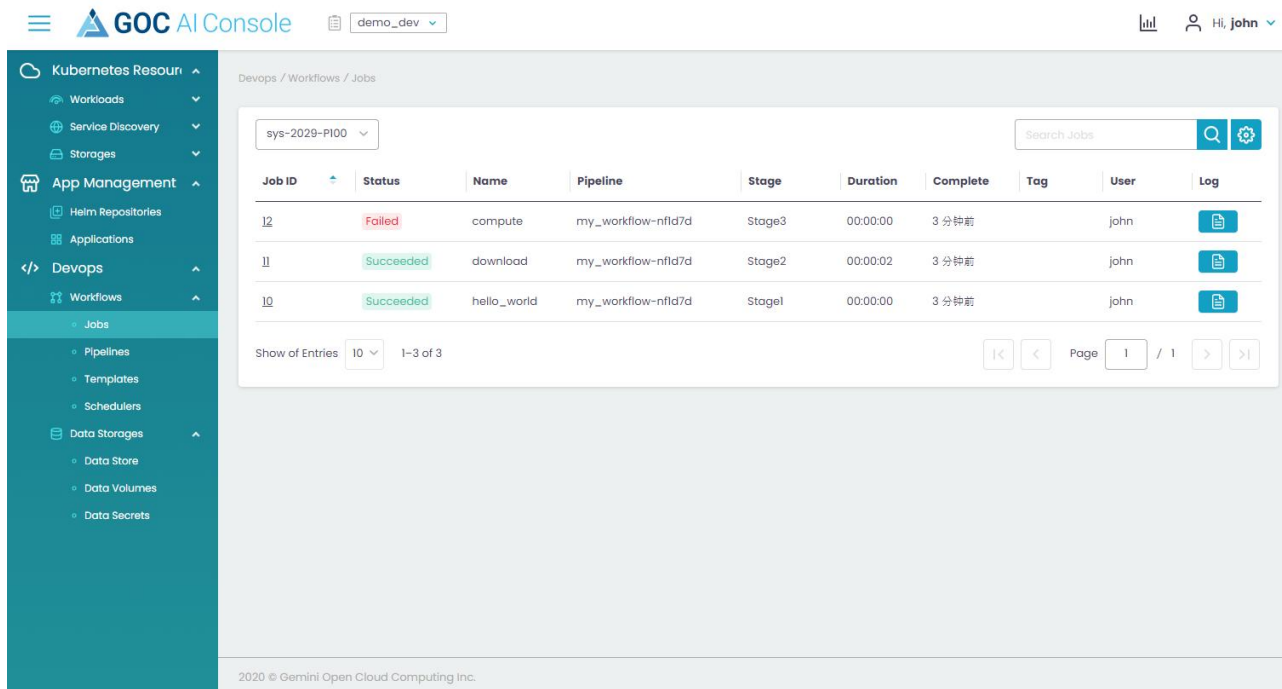
運用GPU共享區資源

透過linux command方式執行
程式




多個Command可以分號『 ; 』
分隔

提供多種GPU數量資源格式運
行容器

任務排程方式執行訓練程式



The screenshot shows the GOC AI Console interface. The left sidebar contains a navigation menu with categories like Kubernetes Resources, App Management, Devops, and Workflows. The main content area displays a table of jobs under the 'Jobs' section. The table has columns for Job ID, Status, Name, Pipeline, Stage, Duration, Complete, Tag, User, and Log. Three jobs are listed: Job ID 12 (Failed), Job ID 11 (Succeeded), and Job ID 10 (Succeeded). Below the table, there are controls for 'Show of Entries' (set to 10) and 'Page' (1 / 1).

Job ID	Status	Name	Pipeline	Stage	Duration	Complete	Tag	User	Log
12	Failed	compute	my_workflow-nfld7d	Stage3	00:00:00	3 分钟前		john	
11	Succeeded	download	my_workflow-nfld7d	Stage2	00:00:02	3 分钟前		john	
10	Succeeded	hello_world	my_workflow-nfld7d	Stage1	00:00:00	3 分钟前		john	

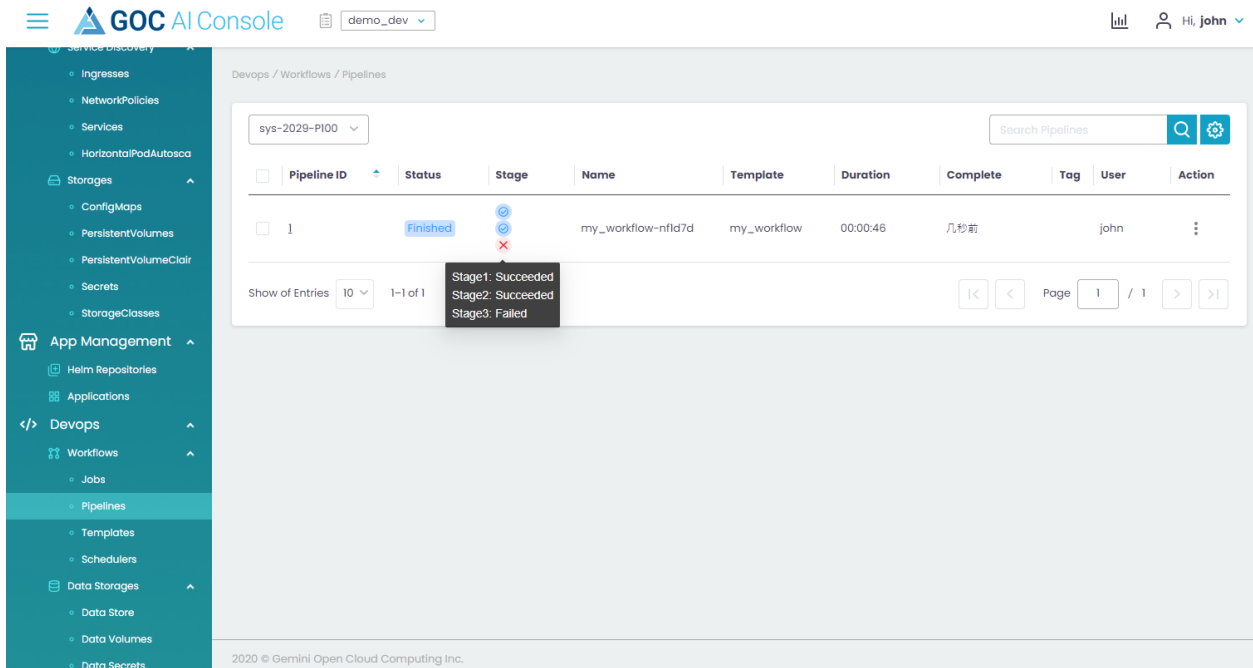
Job延伸-Pipeline(指令多工)

可將多個AI運算工作分成不同階段去執行，當一個階段完成後才會換下一階段去執行，等於將AI運算處理流程化

簡單概念：

單一Job是將單一指令下去運行
Pipeline則是將多個指令分為不同階段下去運行

使用情境：若一個AI訓練透過Tensorflow完成，而我需要用Pytorch或其他工具去驗證結果時，便可使用Pipeline完成。



The screenshot shows the GOC AI Console interface. The left sidebar contains a navigation menu with categories like Service Discovery, Storages, App Management, Devops, and Data Storages. The main content area displays a table of Pipelines. A tooltip is visible over the 'Stage' column of a pipeline, showing the following details:

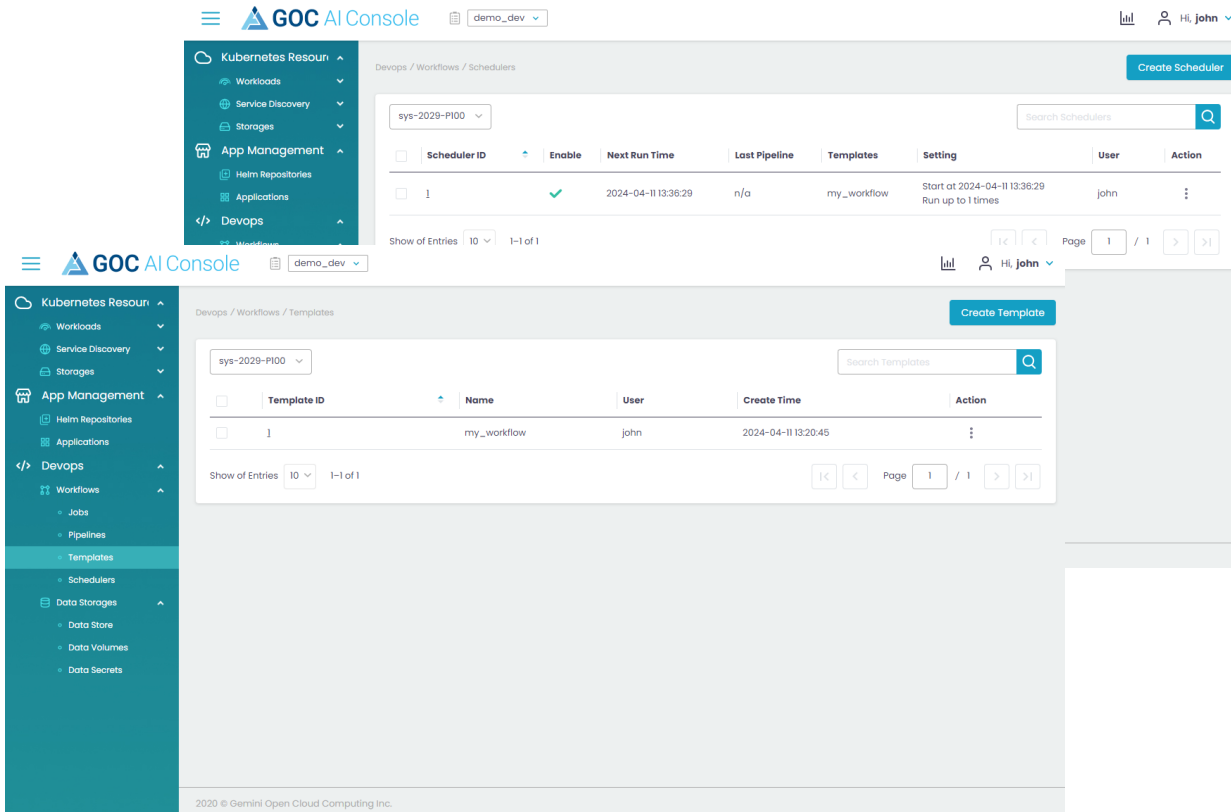
Pipeline ID	Status	Stage	Name	Template	Duration	Complete	Tag	User	Action
1	Finished	Stage1: Succeeded Stage2: Succeeded Stage3: Failed	my_workflow-nfld7d	my_workflow	00:00:46	几秒钟		john	

At the bottom of the console, there is a footer: 2020 © Gemini Open Cloud Computing Inc.

Pipeline 延伸-Template & Scheduler

Template：可將Pipeline儲存成一個固定範本，當有新的資料需要用同樣的運算流程去跑時，只要確保程式碼檔案與資料路徑都正確，就能使用此功能直接運行。

Scheduler：若資料是固定時間會自動放入，則可將現有的Template設定排程，讓AI雲平台能夠在指定的時間自動執行運算。



The screenshot displays the GOC AI Console interface, divided into two panels. The top panel shows the 'Schedulers' management page, and the bottom panel shows the 'Templates' management page. Both panels feature a sidebar with navigation options and a main content area with a table of items and a 'Create' button.

Schedulers Panel:

- Navigation: Kubernetes Resour. > Workloads > Service Discovery > Storages > App Management > Helm Repositories > Applications > Devops
- Table Columns: Scheduler ID, Enable, Next Run Time, Last Pipeline, Templates, Setting, User, Action
- Table Data:

Scheduler ID	Enable	Next Run Time	Last Pipeline	Templates	Setting	User	Action
1	✓	2024-04-11 13:38:29	n/a	my_workflow	Start at 2024-04-11 13:38:29 Run up to 1 times	john	⋮
- Buttons: Create Scheduler, Search Schedulers, Show of Entries (10), Page (1 / 1)

Templates Panel:

- Navigation: Kubernetes Resour. > Workloads > Service Discovery > Storages > App Management > Helm Repositories > Applications > Devops > Workflows > Jobs > Pipelines > Templates > Schedulers > Data Storages > Data Store > Data Volumes > Data Secrets
- Table Columns: Template ID, Name, User, Create Time, Action
- Table Data:

Template ID	Name	User	Create Time	Action
1	my_workflow	john	2024-04-11 13:20:45	⋮
- Buttons: Create Template, Search Templates, Show of Entries (10), Page (1 / 1)

2020 © Gemini Open Cloud Computing Inc.

Hands On

AI雲平台實際操作說明

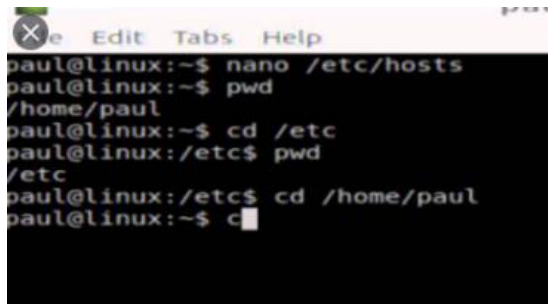
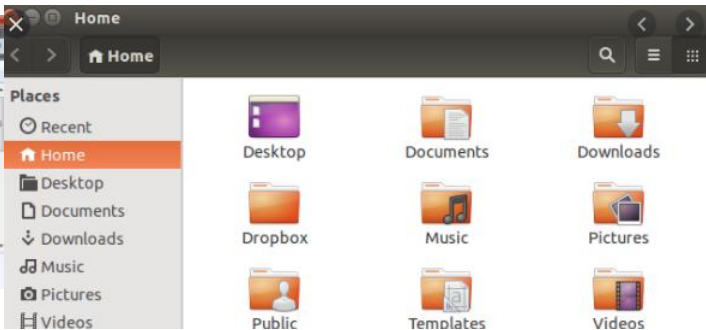
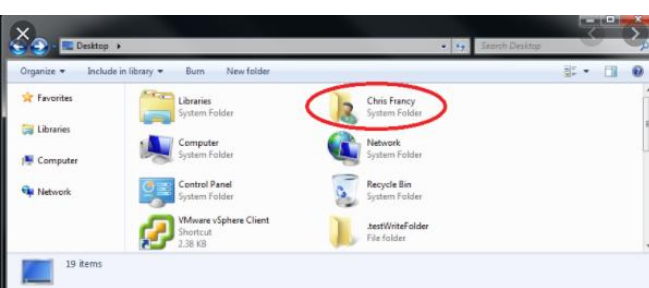
- 1.基本概念：家目錄(home folder)
- 2.Job, Site, User如何存取家目錄
- 3.AI雲平台提供的網頁式編程工具
- 4.如何透過AI雲平台建立Application
- 5.如何透過AI雲平台建立Job
- 6.如何使用沒有GPU的Jupyternotebook環境，動態取用GPU來執行運算

在AI Console專案內，個人資料夾如何使用

1. admin或project admin需要為kubernetes cluster建立StorageClass，讓使用者得知Storage資訊(server, share)
2. User建立第一個PersistentVolumeClaim，以便取得Storage資料夾路徑 (example: “volumeName”:**”pvc-dc5aedf7-14b5-4bdb-ad90-91ac988340ef“**)
3. User在**Devops**建立DataStore與Data Volumes，以便使用相同路徑存取相同資料。
 1. Add Data Store:
 1. NFS Server IP輸入步驟一取得的server ip。
 2. NFS Path輸入share path以及步驟二取得的pvc-xxx合併為同一個NFS path，ex: /nfs/path/pvc-dc5aedf7-14b5-4bdb-ad90-91ac988340ef。
 3. 勾選Using existing NFS data for volume creation
 2. Create Data Volume
 1. Data Volume Setting的地方，Default Container Mount Directory請輸入/home
 2. 勾選Default Data Volume

基本概念：家目錄

- 這是AI雲平台提供給使用者的個人專屬空間。
- 類似Google drive、OneDrive的雲端空間
- AI雲平台的Application, Job透過家目錄，將資料儲存位置串連起來
(類似網路磁碟機的概念)



Job, Site, User如何存取家目錄：1) 透過FTP or SFTP參考管理員提供之方式

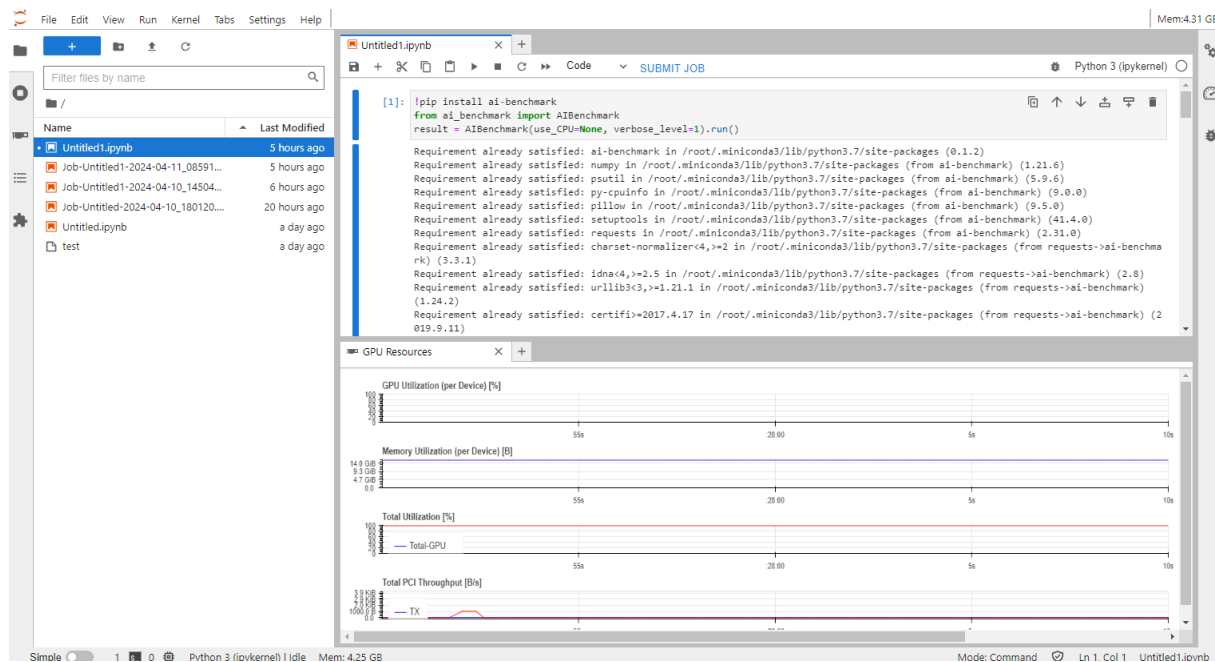
儲存空間伺服器IP:
請參考App Detail

可透過winscp
或
FileZilla等工具連線

Name	torch-site	Endpoints	torch-site-pytorch-jupyter	140.125.248.15:32297 (web)
Status Reason				140.125.248.15:31854 (ssh)
Project	id	3		140.125.248.15:30492 (tensorboard)
	name	test		

Job, Site, User如何存取家目錄：2) 透過Application(常駐型容器)

Application(常駐型容器)裡，
JupyterLab檔案總管預設路徑
(/home)就是使用者資料夾



The screenshot displays the JupyterLab interface. On the left, the file manager shows the home directory (/home) with files like 'Untitled1.ipynb' and 'test'. The main area shows a code cell with the following Python code:

```
[1]: !pip install ai-benchmark
from ai_benchmark import AIBenchmark
result = AIBenchmark(use_CPU=None, verbose_level=1).run()
```

The output of the code cell shows the installation of 'ai-benchmark' and its dependencies, including numpy, psutil, py-cpuinfo, pillow, setuptools, requests, charset-normalizer, idna, urllib3, and certifi. Below the code cell, the 'GPU Resources' section displays four graphs: GPU Utilization (per Device) [%], Memory Utilization (per Device) [B], Total Utilization [%], and Total PCI Throughput [B/s]. The status bar at the bottom indicates 'Python 3 (pykernel) | idle' and 'Mem: 4.25 GB'.

AI雲平台提供的網頁式編程工具

- Jupyter notebook + tensorflow
- Jupyter notebook + pytorch

如何透過AI雲平台建立Application(常駐型容器)

建立Application (常駐型容器)可以快速的取得網頁式的編程工具

- 1.使用帳號登入Portal
- 2.左側功能清單App Management -> Applications，點擊Create Application
- 3.選擇工具類型->輸入資訊->建立
- 4.建立完成後點選該項目名稱->下方顯示的Detail項目
- 5.Endpoints項目清單中，複製(web)該項的IP:Port
- 6.開啟新的瀏覽器，貼上IP:Port
- 6.登入編程工具網頁服務(例如：jupyterlab)

如何透過AI雲平台建立Job(任務型容器)

建立Job(任務型容器)可以取得更多GPU的運算資源來執行程式

使用帳號登入 Portal

左側功能清單Devops -> Workflows -> Templates , 點擊Create Template

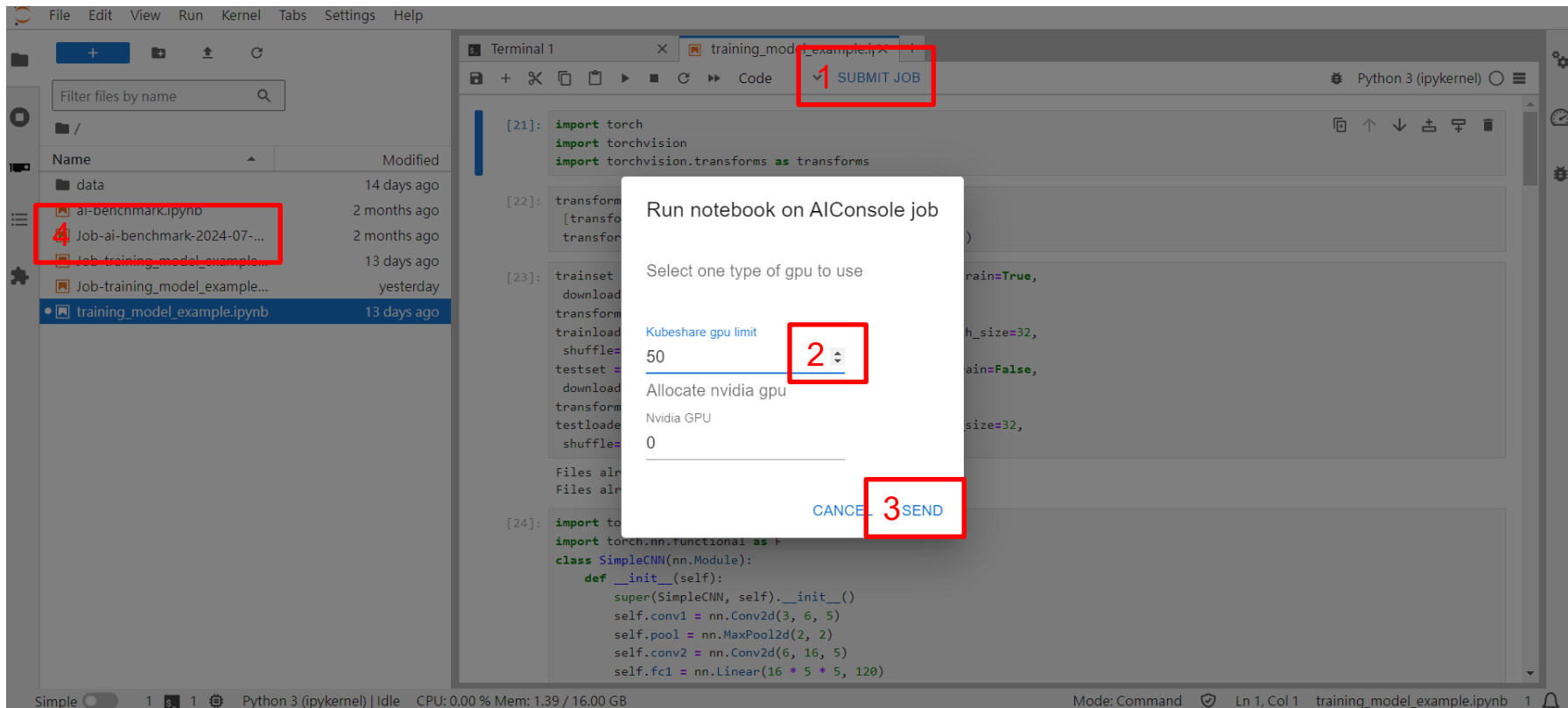
建立任務Template

輸入command

啟動排程(可關閉後再開)

Log查詢執行過程

如何使用沒有GPU的Jupyter Notebook環境，動態取用GPU來執行運算



The screenshot shows the Jupyter Notebook interface with a modal dialog box for running the notebook on an AIConsole job. The dialog box contains the following text and options:

- Run notebook on AIConsole job
- Select one type of gpu to use
- Kubeshare gpu limit: 50 (with a dropdown menu showing '2')
- Allocate nvidia gpu: 0
- Buttons: CANCEL and SEND (with '3' next to SEND)

The background shows the Jupyter Notebook interface with the following elements:

- Terminal 1 window with a 'SUBMIT JOB' button highlighted by a red box and the number '1'.
- Code editor showing Python code for training a model.
- File browser on the left showing a list of files, with 'ai-benchmark.ipynb' and 'Job-ai-benchmark-2024-07-...' highlighted by a red box.
- Bottom status bar showing 'Simple', '1', 'Python 3 (ipykernel) | Idle', 'CPU: 0.00 % Mem: 1.39 / 16.00 GB', 'Mode: Command', 'Ln 1, Col 1', 'training_model_example.ipynb', and '1'.

Q&A