



逢甲大學GPU 加速運算平台 維運經驗分享

逢甲大學資訊處
梁瑞崧、邱清鴻

2021/01/08



大綱

- 目的
- 逢甲大學 GPU 加速運算平台現況
- 建置的歷程
- 遭遇的瓶頸與未來發展的方向



目的

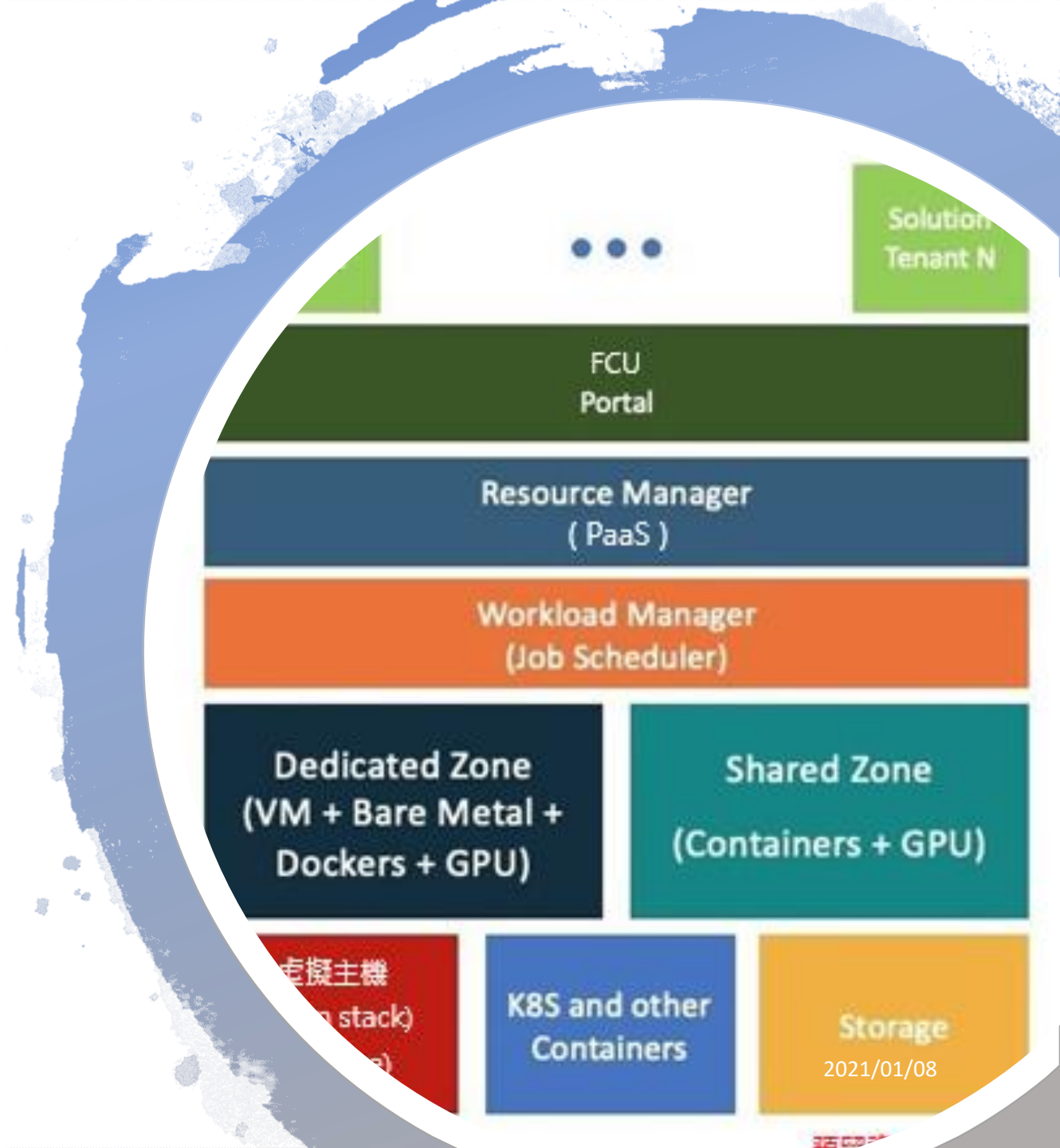
- 近年來因深度學習議題的興起，本校各院系師生開始對相關的應用與議題進行探討，希望能建置一個適用於大部份運算的系統環境，供全校師生使用。
 - 統一的預算支出，獲得較完整的資源。
 - 系統可集中管理且能共享 **GPU** 資源。
 - 簡化使用者維護系統的負擔。



系統規劃方向

- GPU資源可以共享，以申請方式取得資源。
- 統一公版基礎系統，降低系統使用的難度。
- 提供專用的系統環境，使用者可以安裝自己慣用的程式編寫環境，租用時間到期需自行下載運算的結果。
- 提供編寫程式的環境，使用者只能使用制定的程式環境進行運算，可在線上編寫程式，但運算成果需保存於外部儲存空間。

逢甲大學 GPU 加速運算平台現況





系統架構說明

- 專題研究：提供師生專用的作業系統或是程式運算環境。
 - 適用範圍：須要長期重覆調整參數的老師或研究生。
 - 專用的作業系統：以完整的虛擬主機為主，使用者擁有全部的系統權限，作業系統以外的資料保護由使用者自行負責。
- 課程教學：提供程式編寫環境，使用者需具備一定程度的程式編寫和系統操作能力，使用者需自行保護自己的資料。
 - 適用範圍：需要配合課程使用的軟體，以及課後練習。
 - 專用的程式運算環境：以容器化技術提供服務，使用者需具備一定程度的程式編寫和系統操作能力，使用者需自行保護自己的資料。

專題研究

課程教學

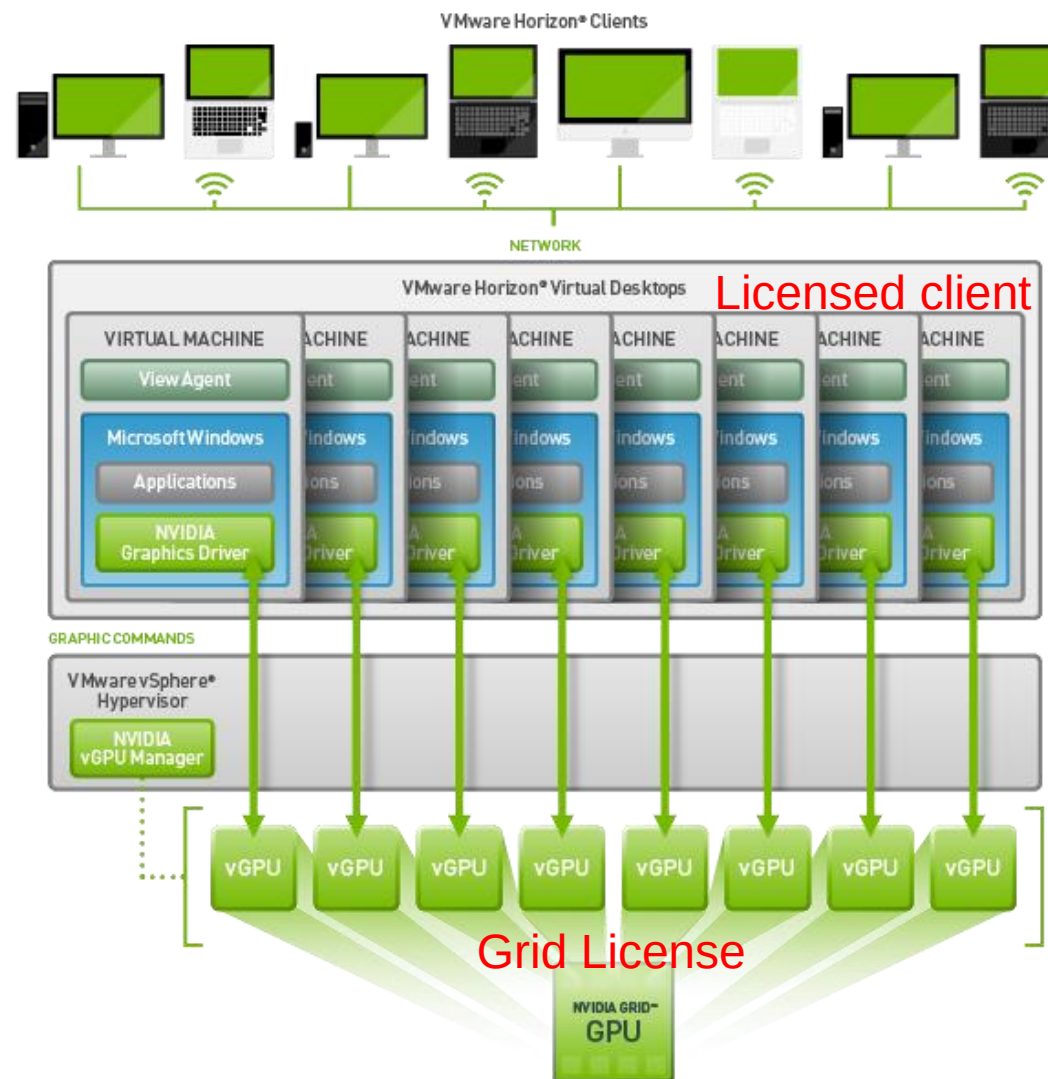
系統登入
(遠端連線、網頁)

虛擬化系統
(VM + GPU)
(K8S + GPU)



專題研究

- 使用 vmware 虛擬化技術提供完整的虛擬主機，透過 nVidia Grid vGPU 技術分割 GPU 資源。
 - 使用者透過電子表單申請使用。
 - 可選擇 8 GB、16 GB、32 GB 和 48G 的 GPU 記憶體。
 - 提供 windows 以及 Linux 作業系統，每台虛擬主機分配 1 個固定 IP，以遠端桌面連線或 SSH 連線方式登入系統使用。





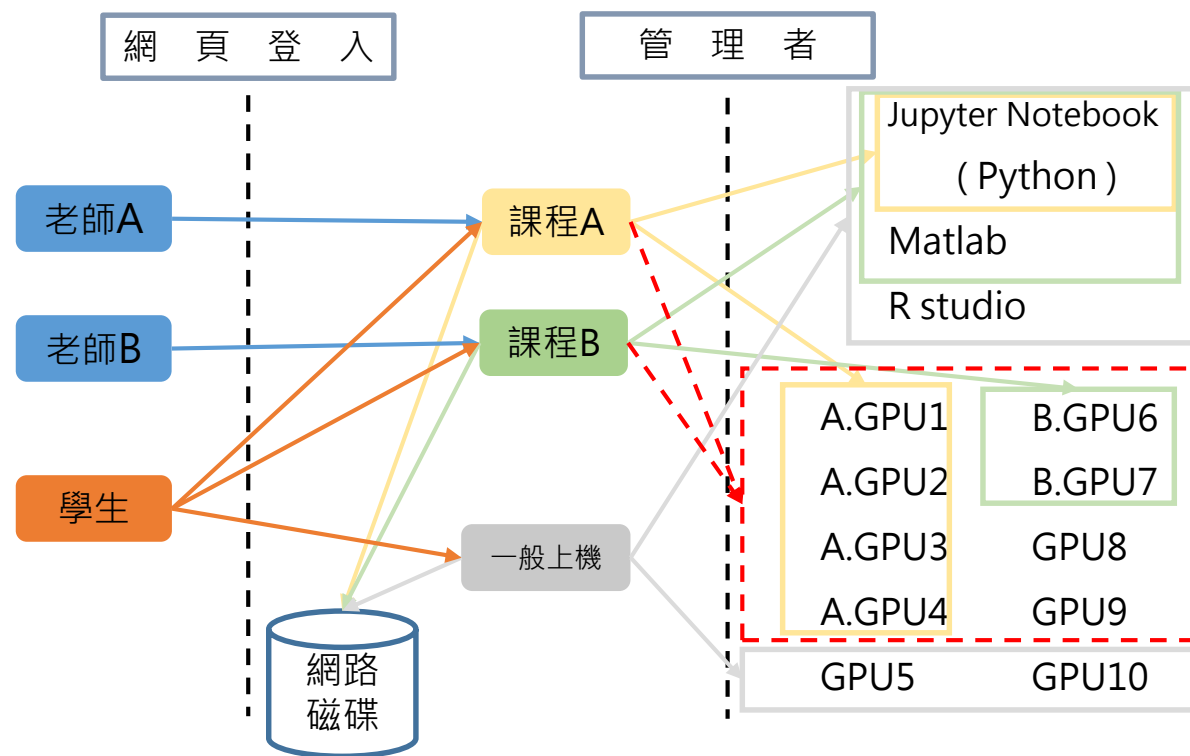
課程教學

- 以容器化技術提供程式編寫環境。

- 使用者可透過申請表單申請課程教學使用。

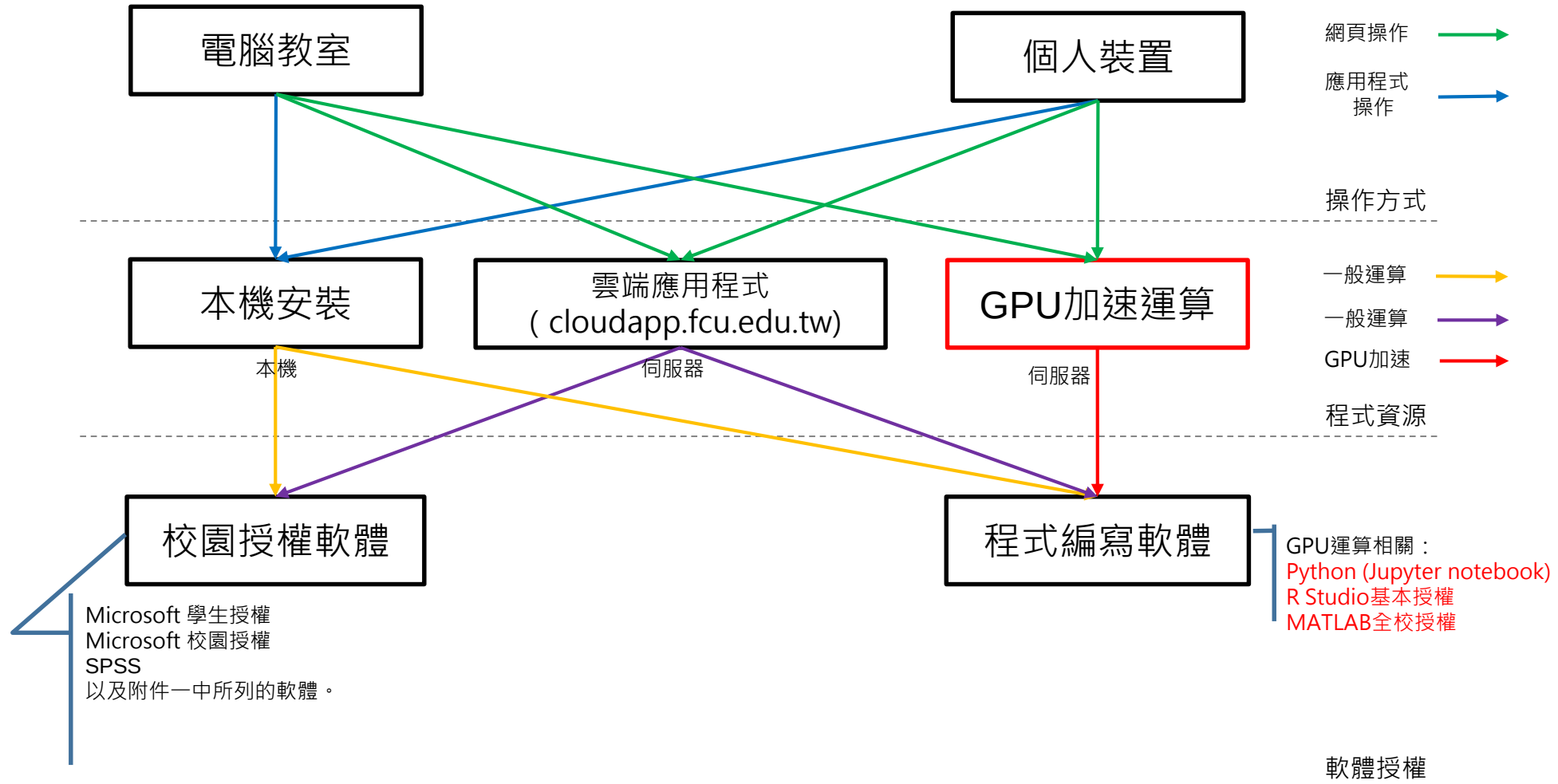
- 視使用現況以及需求調整 GPU 數量。

- 以本校提供的校園帳號，透過網頁入口登入使用，並可掛載外部儲存空間保存資料。





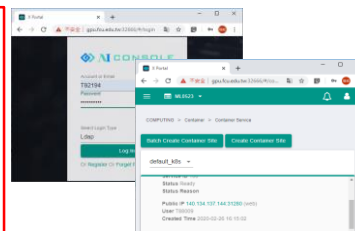
課程教學上課情境示意圖





系統維運方式

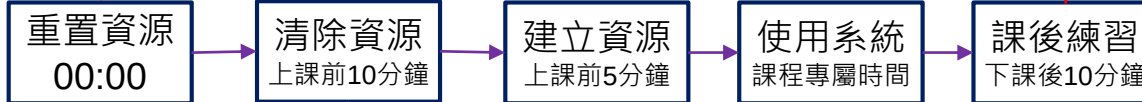
課程教學：使用容器化技術虛擬化應用軟體，使用者**無需自行安裝**校園授權的軟體。



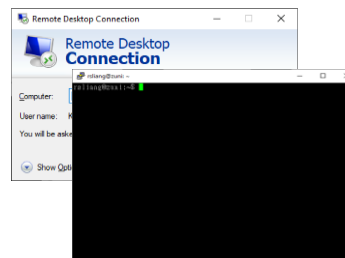
網頁瀏覽器

虛擬應用程式

教學課程
使用時間軸



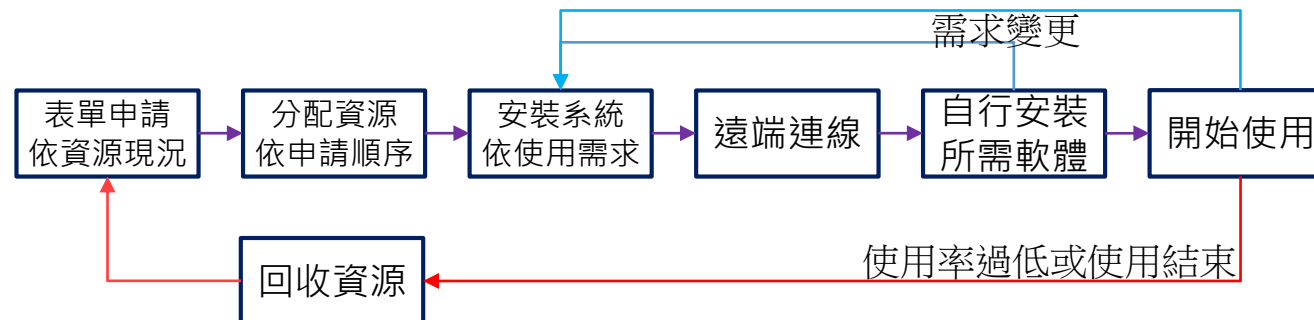
專題研究：提供虛擬主機，使用者可**自行安裝**所需軟體。



遠端連線程式

虛擬主機

專題研究
使用時間軸



虛擬化技術
共用GPU資源

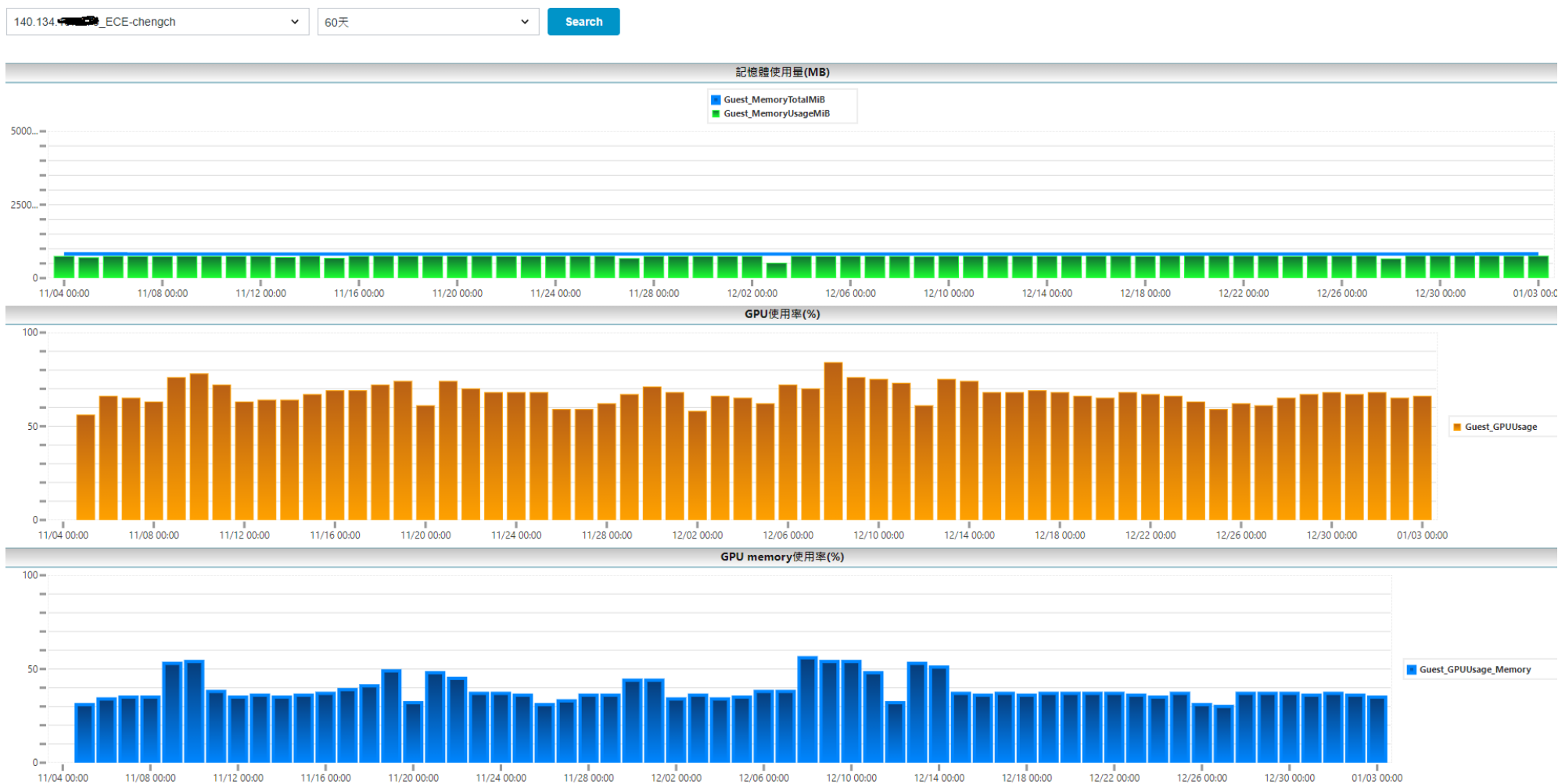


系統資源監控(1/2)

- 以自動排程連線到底層虛擬化系統取得主機和 GPU 資源使用狀態，以補足資源監控數據上的不足。
 - 以SNMP加上 Cacti 保存主機基本的 CPU、記憶體和磁碟的使用狀態，但 SNMP 目前沒有找到合適的方法取得 GPU 的相關資訊。
 - 透過記錄檔分析系統將主機的 CPU、記憶體、磁碟、GPU 運算單元和 GPU 記憶體等的使用狀態做成圖表。



系統資源監控(2/2)



專題研究

課程教

系統登入
(網頁、遠端連線)

虛擬化系統
(K8S : 24 張 GPU)
(VM : 9 張 GPU)

2021/01/08

建置歷程



前期環境評估以及驗證

- 虛擬主機以 vmware horizon 系統共享 GPU，學生可在課堂上和課後可以隨時登入使用 GPU 運算的環境。
- 容器化系統，使用者透過指定的連線 IP 位址登入使用，以人工方式管理。
- 系統環境驗證
 - vmware horizon：成效不如預期，修課人數過多只能分組進行課堂練習，無法有效體驗 GPU 帶來的高速運算效果。
 - 當虛擬主機增加了 GPU 資源後，就再也無法從管理介面進入虛擬主機。
 - 容器化系統：使用者把容器化系統當完整的虛擬主機使用，許多功能在使用上無法滿足使用者需求，且無法驗證快速佈署以及資源再利用的效果。



系統環境調整與使用者回饋(1/2)

- 依用途將 GPU 資源區分為「專題研究」和「課程教學」。
 - 專題研究：以原有 vmware horizon 系統，提供完整的虛擬主機。
 - 課程教學：以 K8S 搭配前台管理程式，提供課程教學專屬環境。
 - 不提供掛載外部儲存空間的功能。
 - 每門課程約 40 位學生。
- 舉辦系統使用以及操作說明會，降低使用者的操作難度。
- 使用者回饋
 - GPU 的記憶體不足，導致沒有節約太多的運算時間。
 - GPU 驅動程式的版本無法依個人要求隨意變更，程式無法直接執行。
 - K8S 系統的操作與傳統的系統操作不同，需有一定程度的使用經驗，才能順利執行運算程式。
 - 掛載外部空間方式需再簡化。
 - GPU數量不足，只能以分組方式進行課程。
 - 容器化系統環境的監測，只在學期末才有 GPU 用量的產生。



系統環境調整與使用者回饋(2/2)

- 依使用者回饋調整系統。
 - 擴充 vmware 系統 GPU 數量。
 - K8S 系統重新調整為「正式區」與「實驗區」。
 - 全校師生可隨時登入系統使用，教學課程在不需使用 GPU 時，可在產生資源時選擇不要分配 GPU 資源。
 - 實驗區每日清除資料，做為程式是否可以在容器化系統中使用的驗證區域，驗證完畢後再將程式移至正式區進行運算。
 - 優化 K8S 系統提供掛載外部儲存空間功能。
- 舉辦使用者經驗分享
- 使用者回饋
 - 提供實驗室以容器化的 Matlab 2020a服務進行影像辨識研究。
 - 多張 GPU 卡同時運算可有效加速運算，運算時間從 30 天左右降低至 14天左右。
 - Matlab 程式介面易當掉。

專題研究

課程教

系統登入
(網頁、遠端連線)

虛擬化系統
(K8S : 24 張 GPU)
(VM : 9 張 GPU)

遭遇的瓶頸與未來
發展的方向



遭遇的瓶頸(1/2)

- 因使用虛擬化技術而衍生的 GPU 驅動程式版本問題。
 - 不管是在虛擬主機還是容器化服務的環境中，使用者必須自行修正程式中與 GPU 有關的函式庫演算法或限制，以符合現有的版本。
- GPU 資源永遠不夠用。
 - 能夠動態調整 GPU 資源增加 GPU 使用率，會比無止盡的跟上新硬體來的有效。
 - 預計以市面上現有的 GPU 共享方式解決，目前皆需改變使用者程式呼叫 GPU 進行運算的習慣。
 - 老師習慣把資源放在實驗室，有需要時可隨時更新或使用，不需要做任何申請的動作。
- 儲存空間存取效能的問題，導致運算效率降低。
 - 改用 SSD，並增加磁碟叢集中的磁碟數量。
 - 提供 LAN 端的 NFS 儲存空間，降低網路的延遲。



遭遇的瓶頸(2/2)

- 虛擬主機當掉，只能人工重啟。
 - 因自動重啟可能會遇到 GPU 自動分派而無法取得應有的資源，此時若無人工介入處理就會無法使用，人工通知重啟反而可以節省時間。
- 容器服務當掉，未存檔的資料就消失了。
 - 程式加入中繼點，減少重新計算時間。
- 使用者需具備一定程度的程式編寫能力。
 - 定期舉辦校內的經驗分享會。
 - 減少有興趣使用的新使用者熟悉系統的時間。
 - 協助使用者瞭解自己需要的是哪一種系統環境。



未來發展的方向(1/4)

- 專題研究將導入 VMware bitfusion 。
 - 改以限制使用上限的方式分配 GPU ，增加使用彈性。
 - 使用者在程式需要 GPU 時才進行呼叫，執行完畢 GPU 歸還給系統。
 - 系統環境驗證
 - 目前 bitfusion 環境只能佈署在 Linux 系統上。
 - 使用 bitfusion 方式呼叫 GPU 資源和直接給定 GPU 資源，在運算的效能上相差不多，但 bitfusion 更具彈性。
 - Toolkit 中的 timeout 參數無法正常回傳數值。
 - bitfusion 的資源監控只能透過管理介面匯出，若需要定期產生相關報表，需再提供相關的 API 。



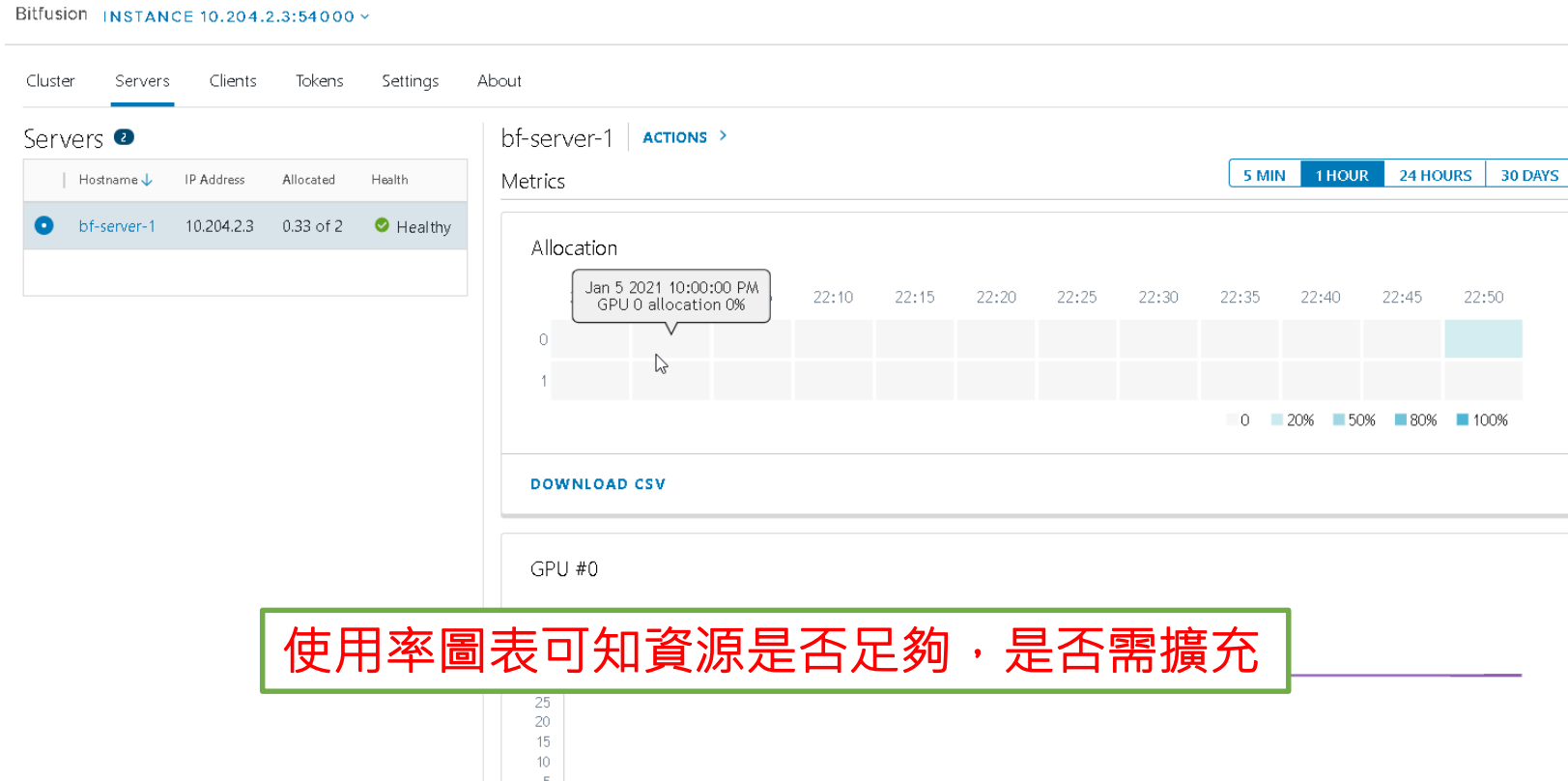
未來發展的方向(2/4)

- 課程教學將導入以下機制以優化系統操作環境。
 - 容器化系統的 GPU sharing 機制。
 - 在使用者多於 GPU 數量時，仍可產生運算資源。
 - 使用者要修改程式，當有閒置資源時，才呼叫 GPU 進行運算。
 - 評估容器化系統的服務狀態備份機制。
 - 降低容器服務異常時資料遺失的風險。
 - 若無適當的解決方案，則使用現有的插入中繼點方式處理。



未來發展的方向 (3/4)

- 優化自動化系統資源監控機制
 - Vmware bitfusion

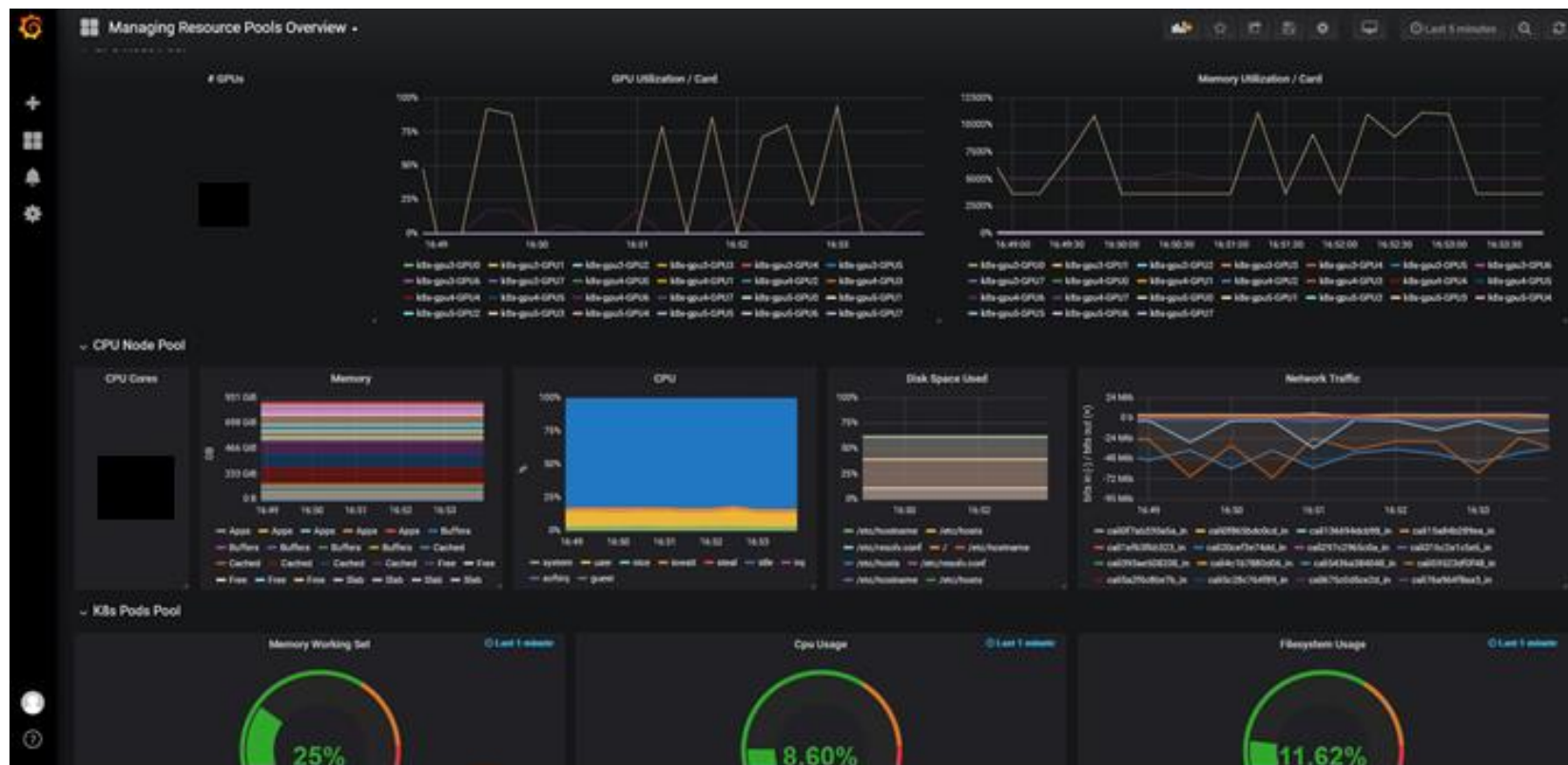


使用率圖表可知資源是否足夠，是否需擴充



未來發展的方向 (4/4)

- 優化自動化系統資源監控機制
 - K8S





- 感謝您的聆聽