



PrimeHub Deploy

企業級 AI / ML 模型部署平台 十倍加速 AI 模型落地運行

概觀

- 提供使用者以瀏覽器(Browser)視覺化介面操作方式，進行模型自助 AI / ML 模型佈署管理作業，既有授權使用者不需要系統管理者協助，即可進行 AI / ML 模型佈署及監視作業環境。
- 提供使用者 AI / ML 模型開發人員模型封裝建議流程及方式。
- 提供同一模型以 Scale-out 水平擴展方式進行大量 AI / ML 模型佈署並得到高可用及負載平衡。
- 提供同一模型以無間斷新舊並行方式進行 AI / ML 模型版本更新作業。
- 提供 AI / ML 模型運作時之視覺化狀態及效能數據監視並提供告警通知。
- 可設定每個 AI / ML 模型運作時環境所需的 CPU、Memory、GPU 使用量，並提供閒置系統資源自動回收。
- 可指定使用者群組是否能夠進行 AI / ML 模型佈署作業。
- AI / ML 模型應用端點安全金鑰認證功能。

主要優點

- 快速安裝設定後立刻可用
- 簡單易懂的操作管理介面，相關說明文件齊全
- 運算資源管理: 虛擬切割運算資源，提供多模型共用
- 確實的使用者及團隊權限安全管理
- 資料集權限管理: 可針對使用者及群組讀取寫入管控
- 映像檔管理: 使用者可輕鬆選擇適合的映像檔
- 帳號系統整合 AD / LDAP
- 服務監控瞭解使用狀況，提供即時告警機制
- AI / ML 模型應用端點可透過金鑰安全機制認證

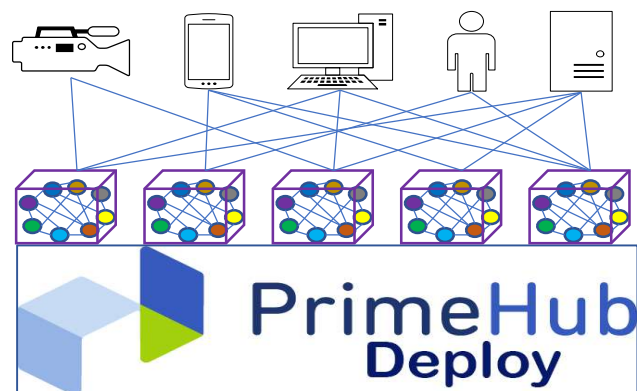
關於 PrimeHub Deploy

PrimeHub Deploy 人工智慧模型之部署運行管理及監控平台，提供流暢精簡的人工智慧模型之部署流程，得以十倍加速對模型加以封裝、推送、佈建、運行及監控實現最佳使用體驗。提供以 API 介接方式，可搭配用戶自建之 CI / CD 流程機制，對人工智慧模型之部署得以自動化管控與稽核。使用以 Kubernetes 為運行基礎平台，可以獲得最大的安全、可靠與彈性，簡單易懂的管理介面大幅降低操作與管理負擔，使得 AI / ML 團隊人員即可自主管理運作。

AI 機器學習或深度學習部門，已經有 AI / ML 模型產出需部署運行維運管理，請立即開始使用 PrimeHub Deploy。

版本說明

版本	說明	授權數量	服務等級
Pre-Production	適用於開發測試環境使用	沒有限制	普通級
Small Production	適用於小型部署環境使用	<=10	企業級
Medium Production	適用於中型部署環境使用	<=50	企業級
Large Production	適用於大型部署環境使用	<=200	企業級
Extra Production	適用於超大型部署環境使用	>201	企業級



詳細資訊

想要了解其他資訊或購買 PrimeHub Deploy 產品，請致電 +886-2-6604-0706 Ext 200、造訪我們的網站 <https://InfuseAi.io>，或聯絡我們的授權經銷商。詳細的產品規格和系統需求，請參閱 [PrimeHub](#) 說明文件。

試用申請

如有需求想要試用 PrimeHub Deploy 請於下列網址填寫申請表格，我們將儘速回應您的需求。
<https://forms.gle/TEwPdb57Nqqkiwrz6>

PrimeHub MLOps Platform

協助 AI 模型開發、管理、部署到監控的完整平台

關於

PrimeHub MLOps 人工智慧模型開發及部署平台為人工智慧開發、部署、管理及監控整合平台，提供流暢精簡的人工智慧模型開發及部署流程，加速人工智慧模型落地，提供開發作業、版本管理、部署運行及監控實現最佳人工智慧整合平台。

提供以 API及SDK 介接方式，可搭配用戶自建之 CI / CD 流程機制或是藉由應用程式內嵌SDK控制，對人工智慧模型之模型訓練及部署得以自動化管控與稽核。

透過APPs 整合第三方軟體，開創無限整合可能

使用以 Kubernetes 為運行基礎平台，可以獲得最大的安全性、可靠性與擴充彈性，簡單易懂的管理介面大幅降低操作與管理負擔，使得 AI / ML 團隊人員即可自主管理運作。

AI 機器學習或深度學習部門想要加速人工智慧模型落地、降低開發管理時間成本，請立即開始使用 PrimeHub MLOps 平台。

主要優點

安裝設定簡易快速，完成後立刻可用

簡單易懂的操作管理介面，相關說明文件齊全

基於 Kubernetes 虛擬化提供安全穩健的系統資源分配

確實的使用者及團隊權限安全管控

資料集權限管理可針對使用者及群組讀取寫入管控

API、SDK及CLI整合既有軟體系統實現自動化

映像檔管理可輕鬆選擇適合的人工智慧開發環境

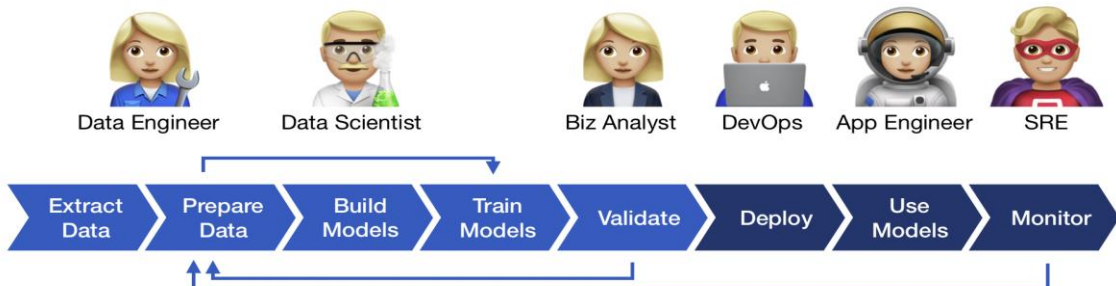
APPs整合第三方軟體豐富MLOps整合

模型版本管理詳實記錄模型開發歷程

帳號系統整合 AD / LDAP 安全有保障

服務監控瞭解使用狀況，提供即時告警機制

AI / ML模型應用端點可透過金鑰安全機制認證



主要模組

- PrimeHub MLOps platform base software kit
- Jupyter notebook
- Job submission
- Job schedule
- Model management
- Model deployment
- Image builder
- Monitoring
- APPs
- User authorize

軟體最低需求

- Kubernetes 1.19+
- PrimeHub MLOps platform requires 10 vCPU · 12GB RAM · 2TB storage
- Model Deploy requires 6 vCPU, 6GB RAM, 100GB storage
- Recommend adding 1 CPU, 2GB RAM per model
- Optional MinIO: Server mode requires 2 CPU, 4GB RAM, 100GB storage. Gateway mode requires 1 CPU, 2GB RAM
- Optional cluster and model monitoring requires additional 4 vCPU, 8GB RAM, 120GB storage

詳細資訊

想要了解其他資訊或購買 PrimeHub MLOps 產品，請致電 +886-2-6604-0706 Ext 200，造訪我們的網站 <https://www.infuseai.io>，或聯絡我們的授權經銷商。詳細的產品規格和系統需求，請參閱 [PrimeHub](#) 說明文件。

試用資訊

如有需求想要試用 PrimeHub MLOps 請於下列網址填寫申請表格，我們將儘速回應您的需求。
<https://www.infuseai.io/request-demo>