KeyWin

從開機到就緒：NVIDIA DGX Spark 快速入門指南

NVIDIA DGX Spark™ 已正式向 AI 開發者供貨，

對於剛入手的全新 DGX Spark，

該如何進行初始化設定？本篇文章將引導您完成 DGX Spark 首次設定。

在初始設定的過程中，您需要選擇登入系統的方式，

並執行首次設定實用程式來配置所有內容。

設定完成後，可以根據喜好選擇不同的方式存取DGX Spark。

本文內容主題：

- DGX Spark 硬體概述
- 開箱並連接設備
- 首次啟動與初始設定
- 系統資訊確認
- SSH 命令列遠端登入
- Docker 配置

一、 DGX Spark 硬體概述

作為新一代電腦，DGX Spark 以小巧的桌面機形態提供 1 PFLOP AI 效能和 128GB 統一記憶體(Unified Memory)，使開發者能夠在本機進行最高 2,000 億參數的 AI 模型推理以及對 700 億參數的模型進行微調。此外，開發者還可通過 DGX Spark 在本機建立 AI代理和進行進階軟體堆疊。

DGX Spark 具有以下特點：

- 搭載 GB10 超級晶片
- 內建 20 核心高效能 Arm 架構 的 NVIDIA Grace CPU
- 128 GB 統一共享記憶體
- 進階連接功能，包括高效能 NVIDIA ConnectX™ 網路技術、Wi-Fi 7 等
- 支援最高 200B 參數等級的 AI 模型（雙 DGX Spark 互聯可支援最高 405B 模型）
- 小巧桌上型外觀

NVIDIA DGX™ Spark Founders Edition

- 搭載 GB10 超級晶片
- 每秒高達 1,000 TOPS 的 AI 效能（FP4 精度）
- 配備第五代 Tensor Core 技術的 NVIDIA Blackwell GPU
- 內建 20 核心高效能 Arm 架構 的 NVIDIA Grace CPU
- 高達 128 GB 統一共享記憶體，運算資料即時同步
- 最多可達至 4TB NVMe 儲存空間
- 支援多達 200B 參數的大型語言模型
- 搭配 NVIDIA ConnectX 網路可串接兩台 DGX Spark，擴展至 405B 參數模型
- 桌上型精巧設計
- 標準插座供電，即插即用

二、 開箱並連接設備

DGX Spark 包含硬體及配件如下：

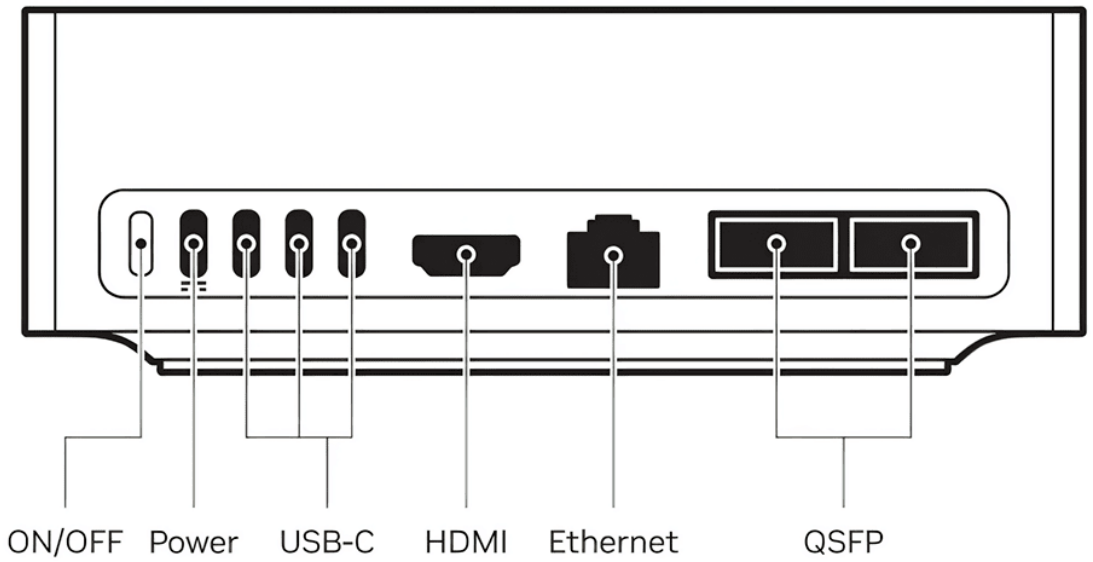
- 1 台 DGX Spark
- 快速入門指南手冊
- 電源轉接器
- 電源線

DGX Spark 支援兩種模式：

- 顯示器模式：連接螢幕、鍵盤、滑鼠
- 遠端模式：透過 SSH 遠端登入

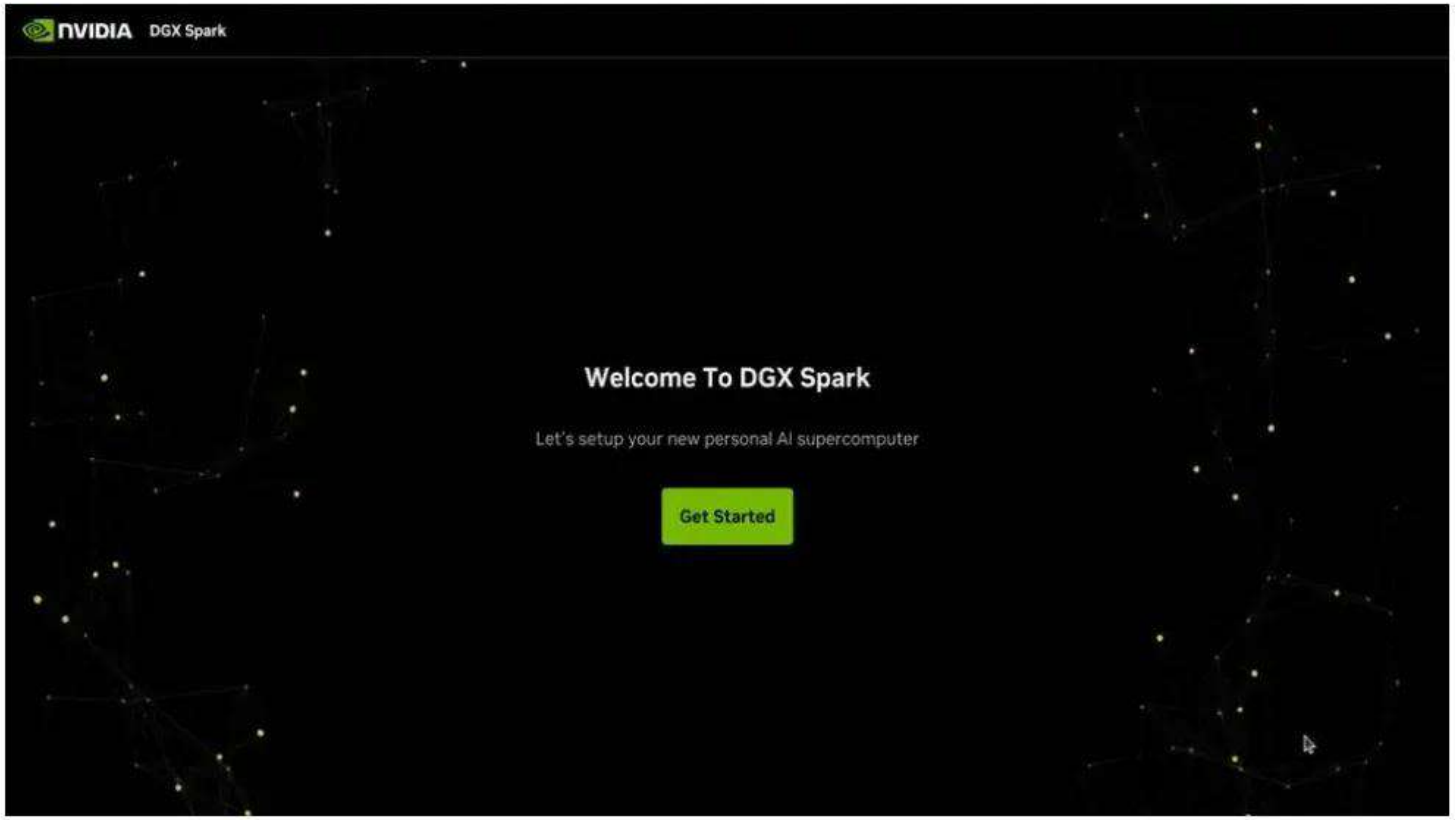
此處示範顯示器模式，依以下順序連接：

- HDMI 螢幕
- USB 鍵盤與滑鼠（若開機後未偵測到鍵盤或滑鼠，系統會提示您將藍牙裝置切換至配對模式）
- 電源供應器（僅能插在最左側 Type-C 連接埠）
- 按電源鍵開機（如下圖最左側按鈕）



三、 首次啟動與初始設定

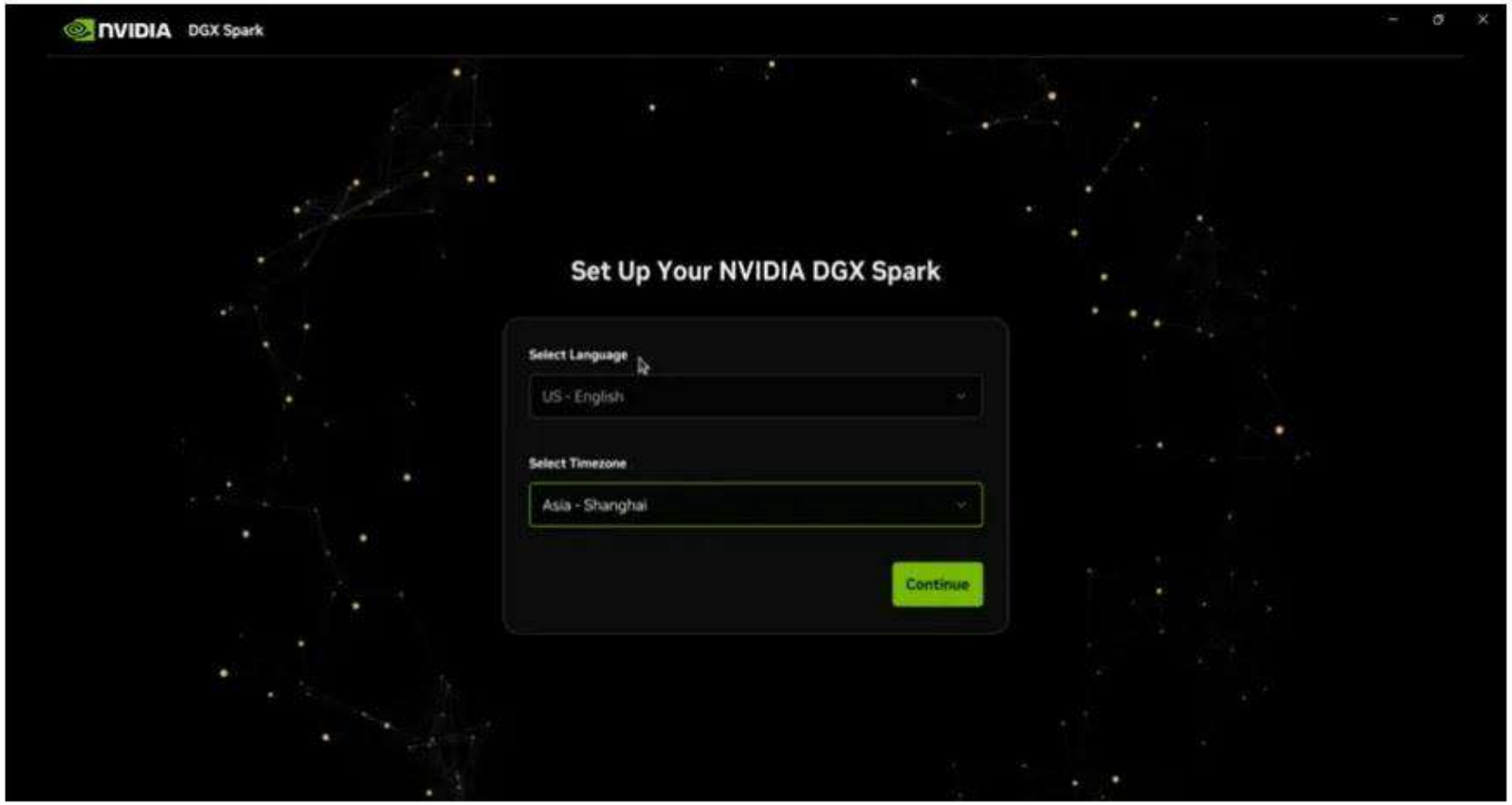
系統啟動後，已連接的螢幕會顯示 DGX Spark 自動載入的首次設定精靈，引導您完成初始化設定。點擊「Get Started」後依畫面提示逐步完成設定。



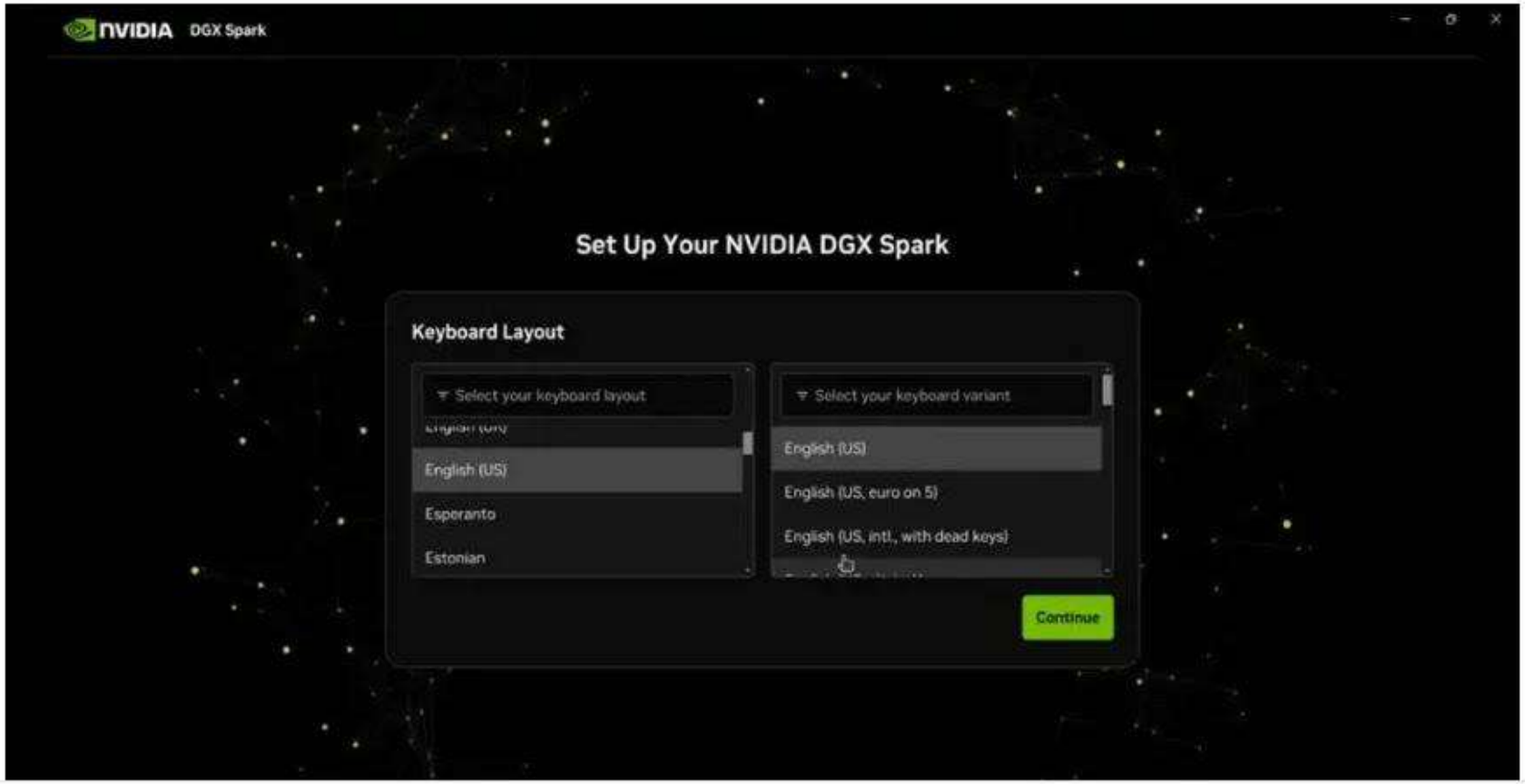
設定流程：

1.語言與時區選擇

選擇偏好的系統語言與時區。 輸入框會隨輸入自動篩選。

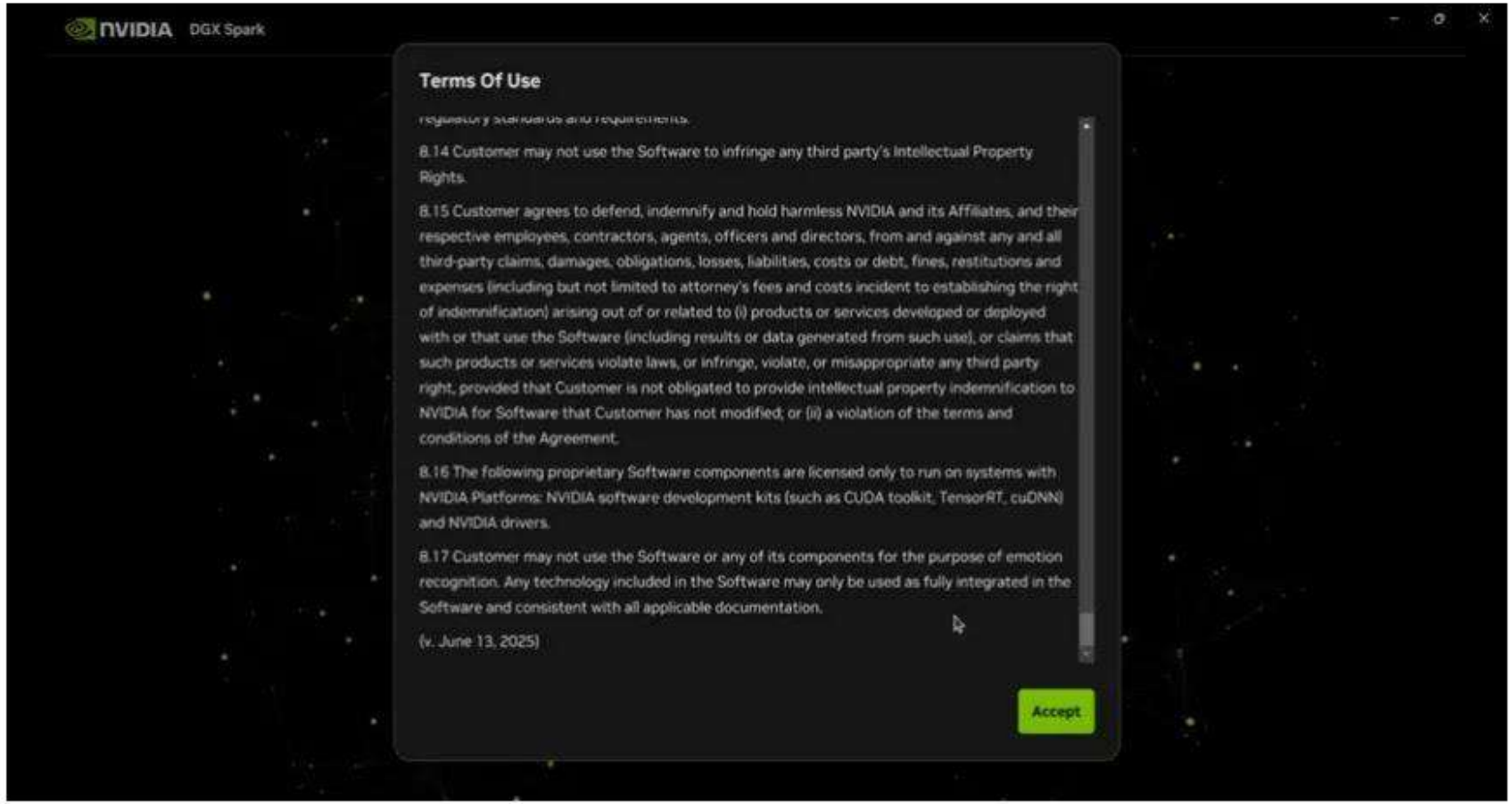


2.鍵盤配置（僅限使用顯示器時）

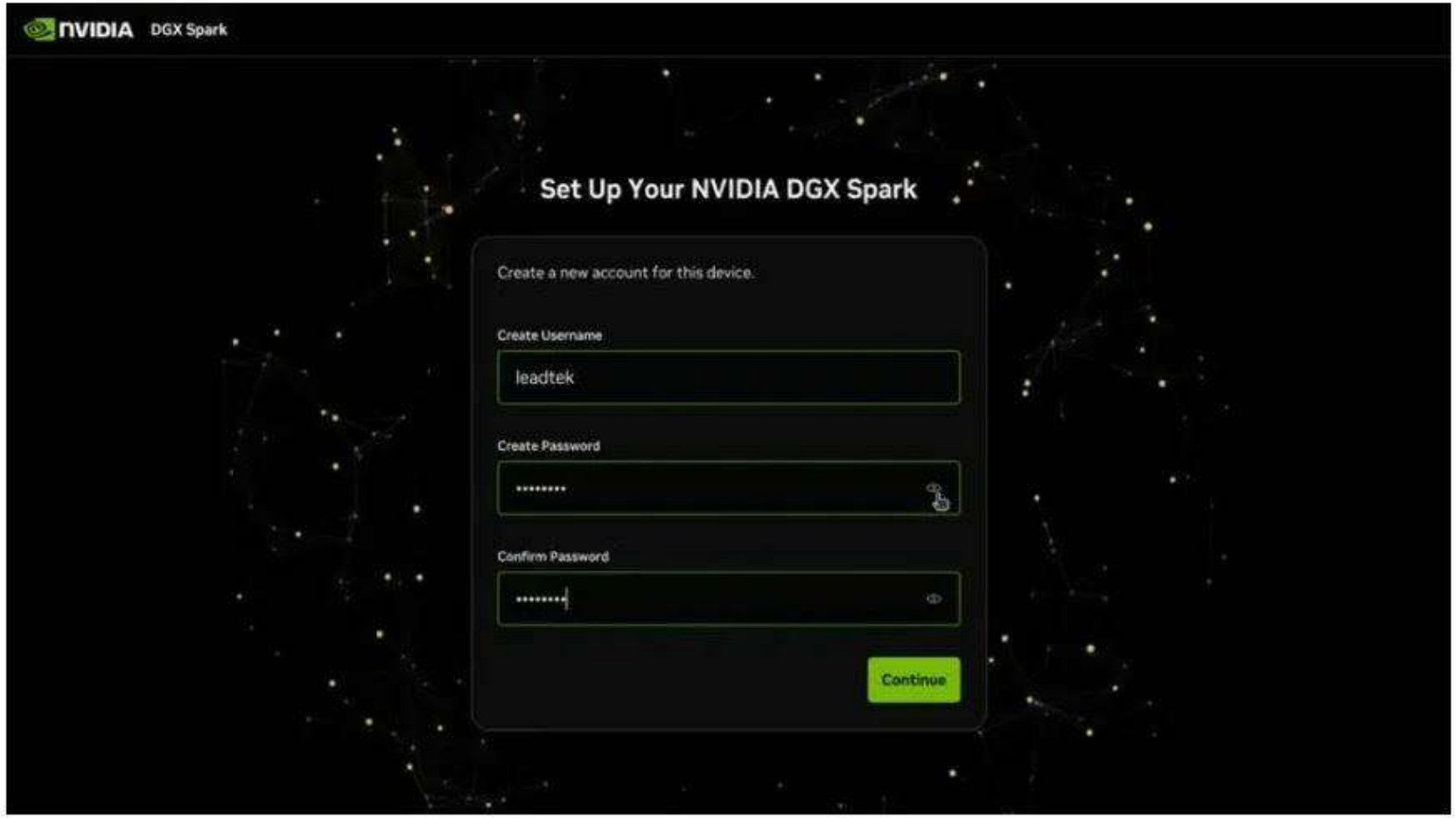


3.使用條款

閱讀並接受後繼續。

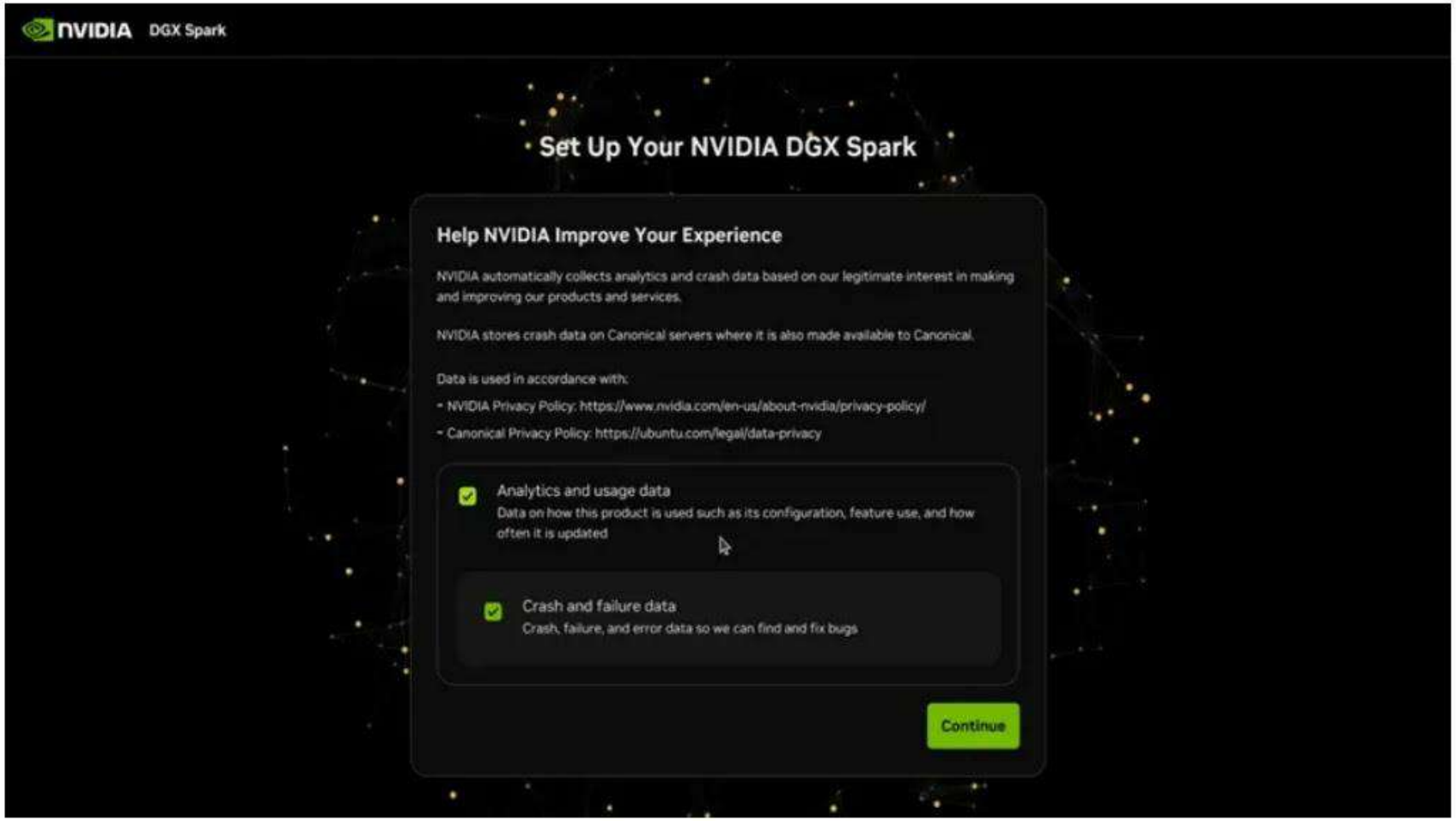


4.建立使用者帳號



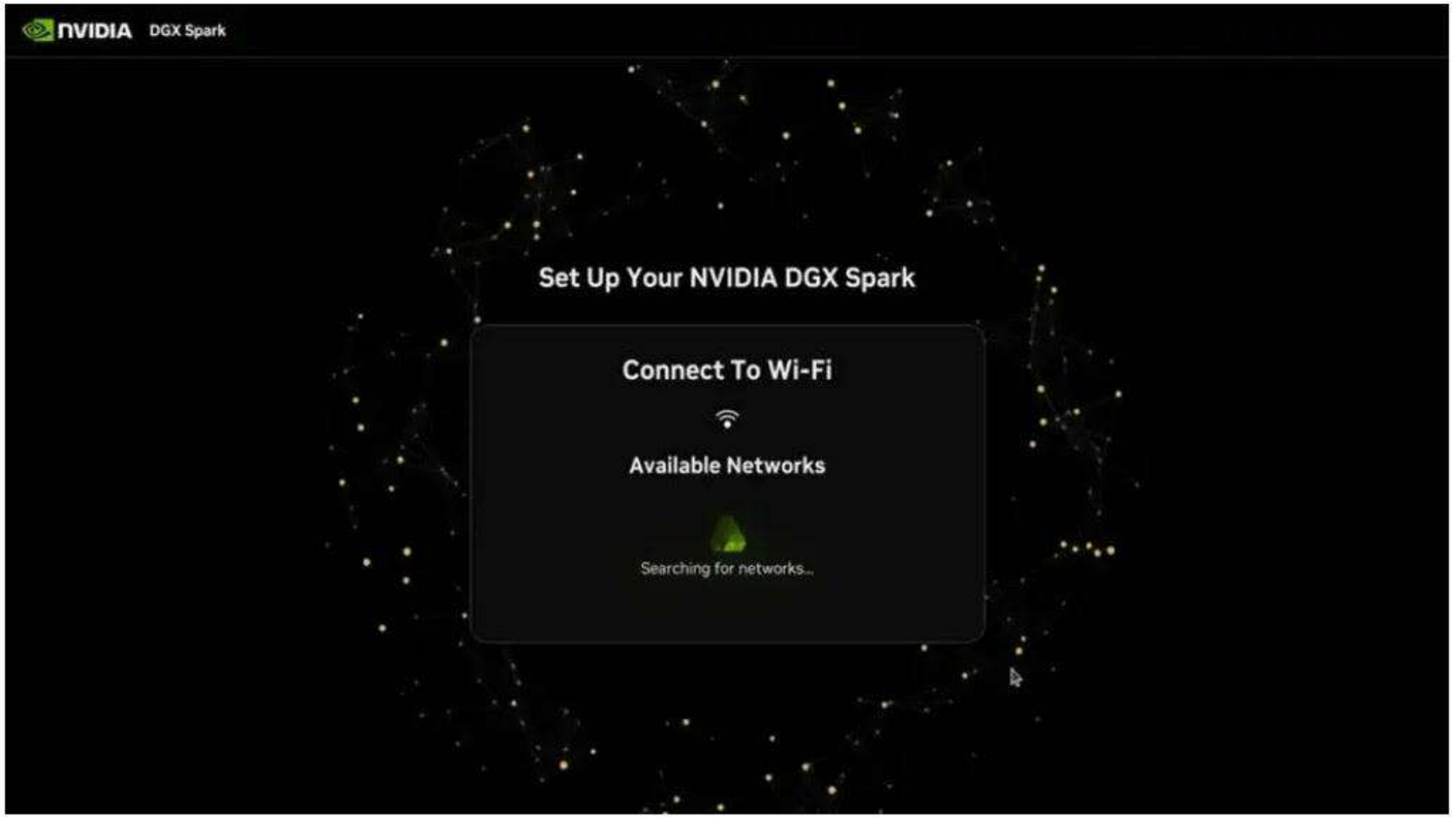
5.資訊分享設定 (選填)

配置分析和錯誤報告設定，此步驟可跳過。



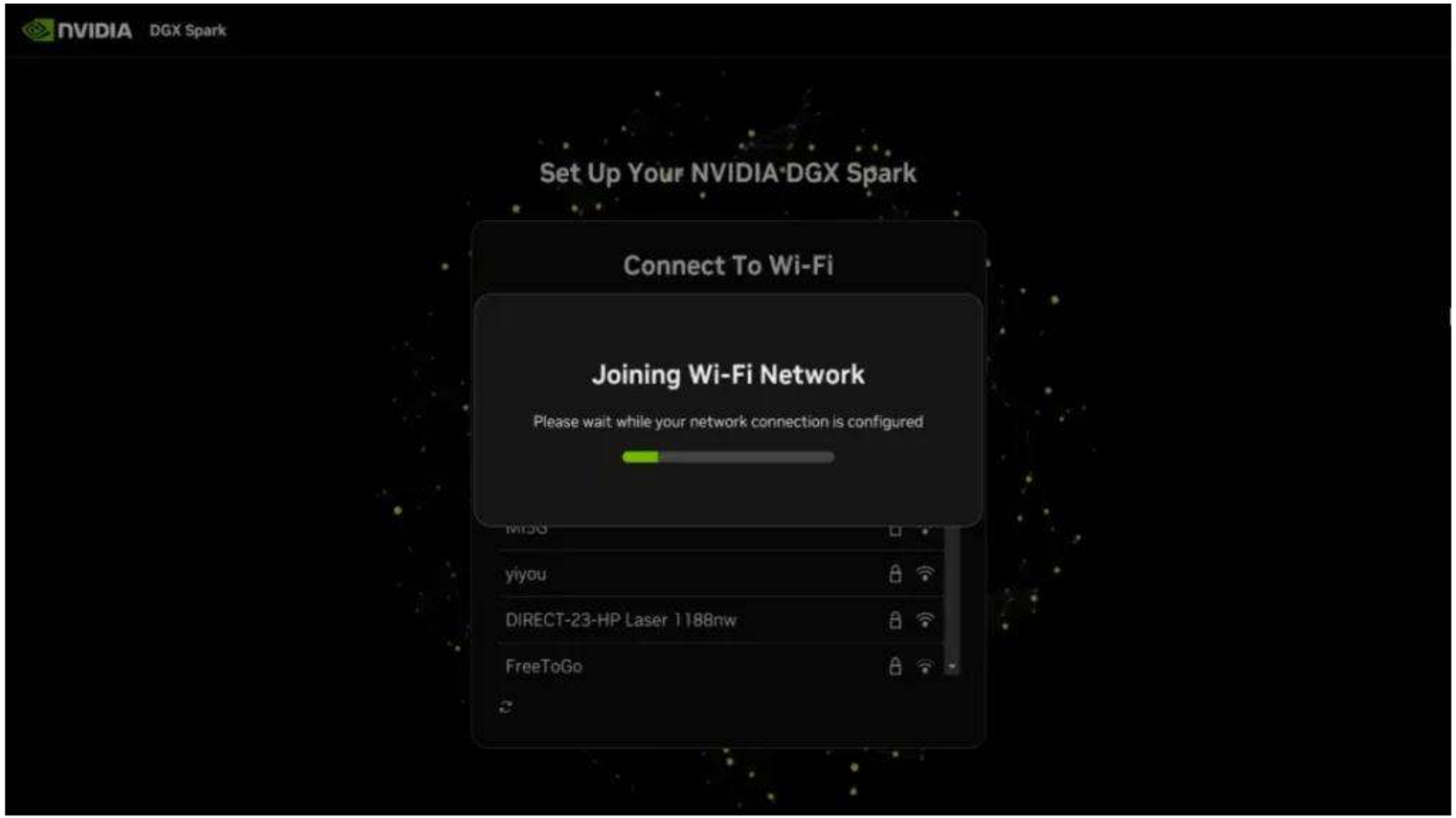
6.Wi-Fi 網路選擇

若已插上有線網路並成功上網，此步驟會跳過。



KeyWin | Server Hardware Supplier 凱穩電腦

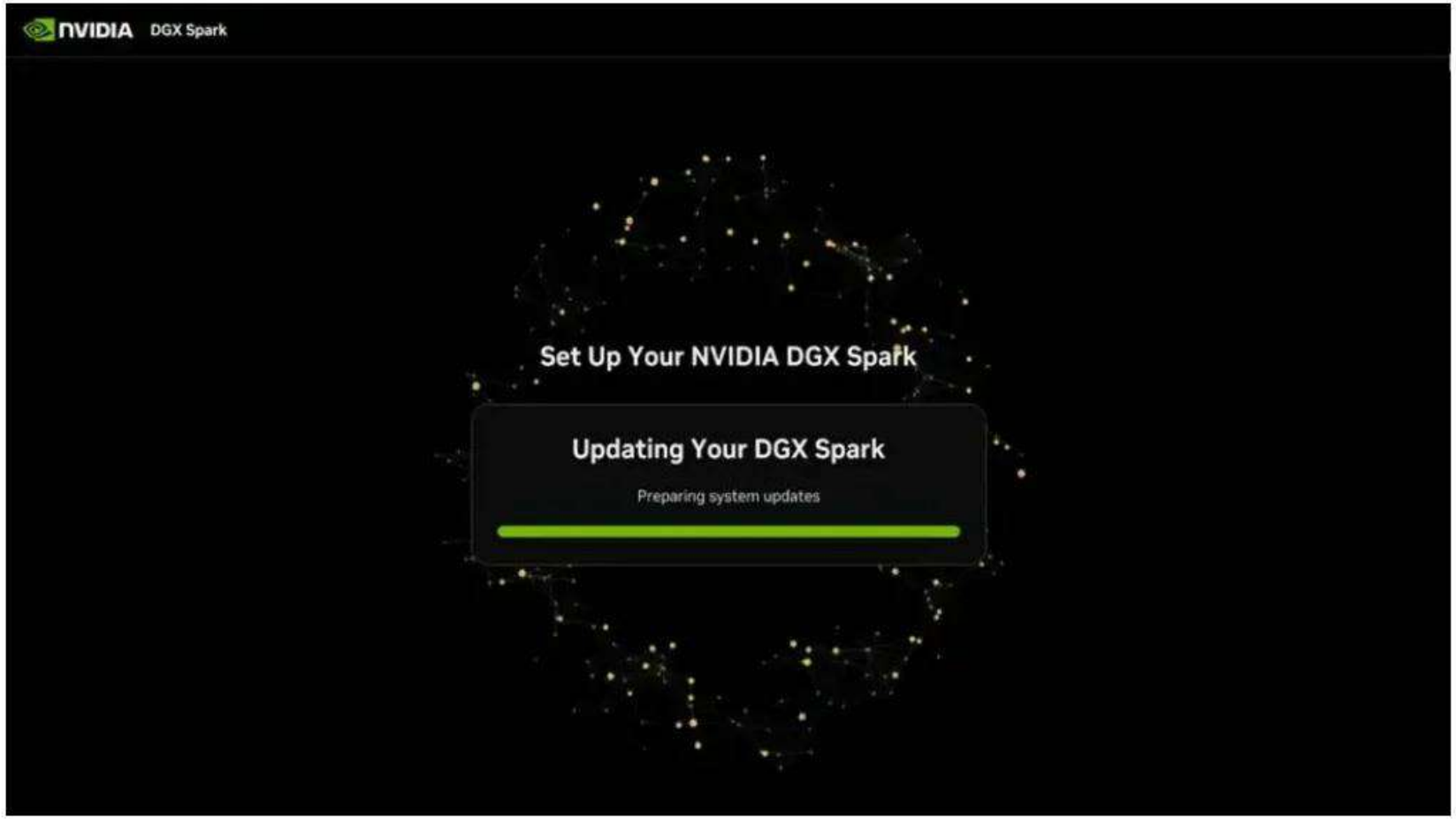
7. Wi-Fi 密碼輸入

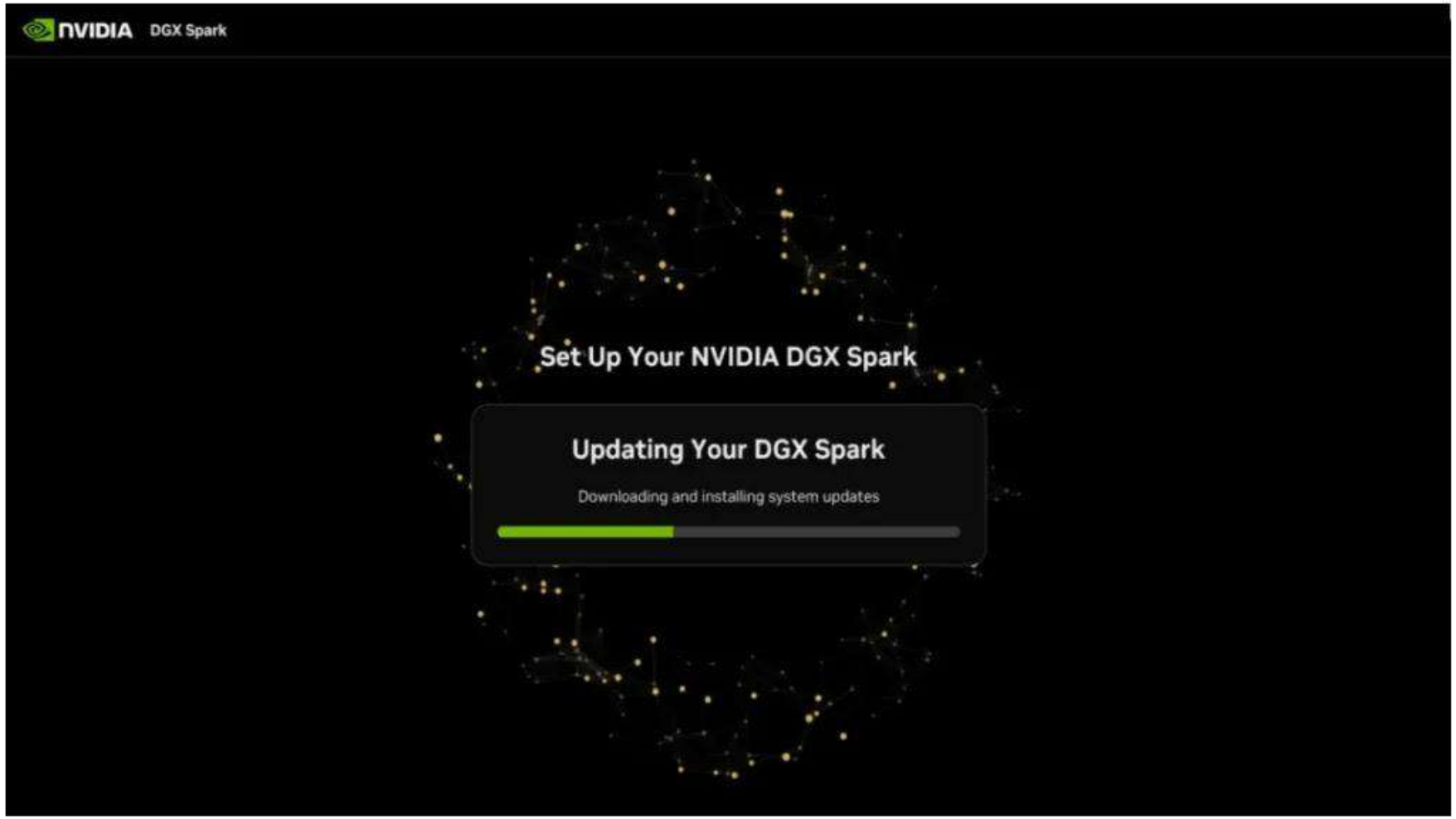


8.軟體下載與安裝

連上網路後系統會自動下載完整軟體並安裝。此過程需要一些時間，可能會重新開機多次。

⚠ 在此過程中請勿關閉或重新開機。下載開始後，安裝無法中斷。





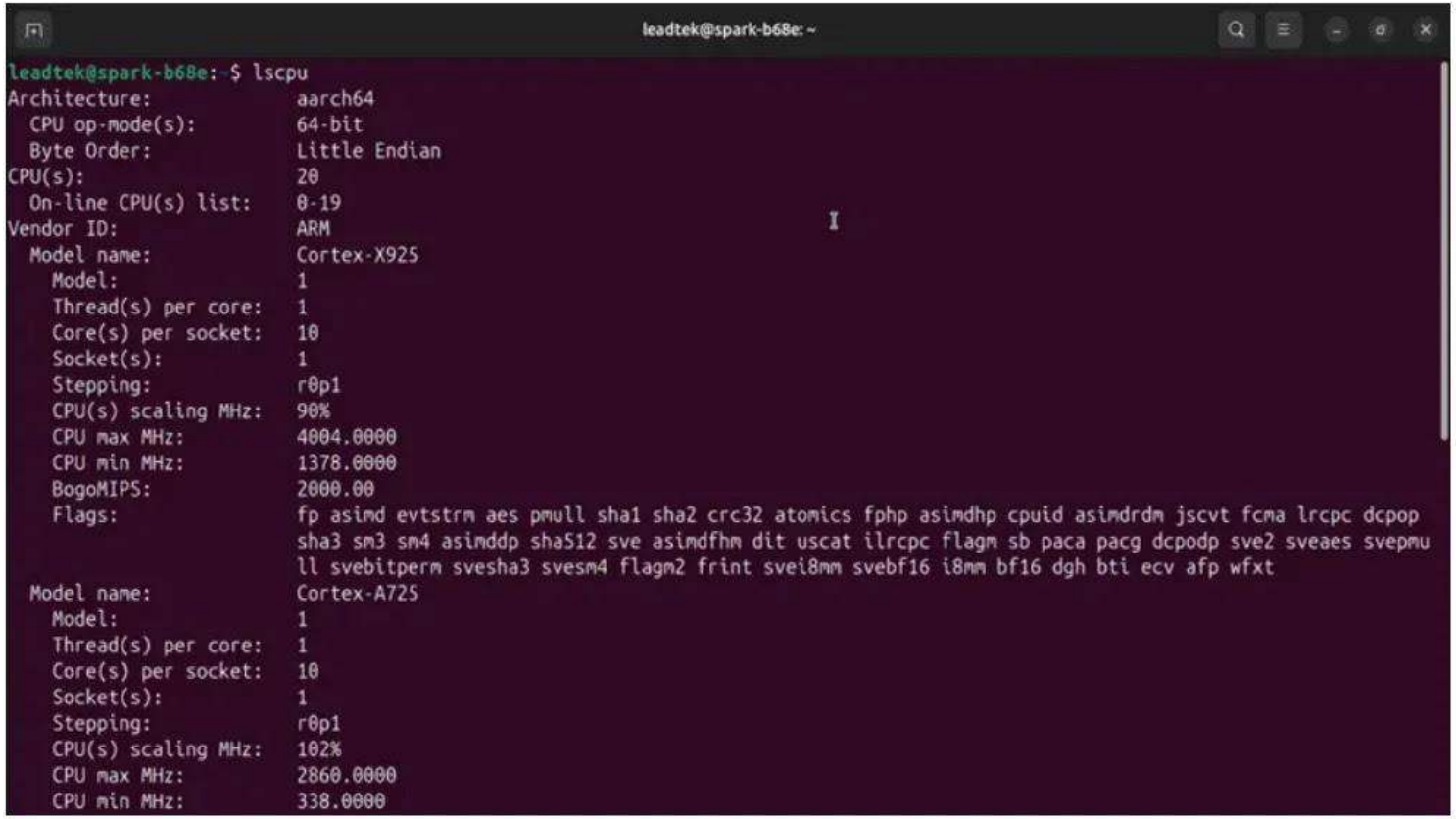
9.安裝完成

安裝完成後，設備會自動重開機，之後即可正常使用。

四、 系統資訊確認

進入桌面後，按下 Ctrl + Alt + T 開啟終端機，用以下指令查看系統資訊：

- lscpu 查看 CPU 資訊



- free -h 查看記憶體容量

```
leadtek@spark-b68e:~$ free -h
               total        used        free      shared  buff/cache   available
Mem:           119Gi       4.4Gi       114Gi        13Mi       1.8Gi       115Gi
Swap:           15Gi          0B        15Gi
```

- lsblk 查看儲存容量

```
leadtek@spark-b68e:~$ lsblk
NAME        MAJ:MIN RM   SIZE RO TYPE MOUNTPOINTS
loop0         7:0    0     4K  1 loop /snap/bare/5
loop1         7:1    0  68.9M  1 loop /snap/core22/2140
loop2         7:2    0 232.8M  1 loop /snap/firefox/7182
loop3         7:3    0  91.7M  1 loop /snap/gtk-common-themes/1535
loop4         7:4    0 493.6M  1 loop /snap/gnome-42-2204/228
loop5         7:5    0 218.5M  1 loop /snap/thunderbird/850
loop6         7:6    0  44.3M  1 loop /snap/snapd/25585
sda           8:0    1     0B  0 disk
sdb           8:16   1     0B  0 disk
nvme0n1     259:0    0   3.7T  0 disk
├─nvme0n1p1 259:1    0   298M  0 part /boot/efi
└─nvme0n1p2 259:2    0   3.7T  0 part /
```

- nvidia-smi 查看 GPU 狀態

```
leadtek@spark-b68e:~$ nvidia-smi
Wed Nov 12 11:34:39 2025
+-----------------------------------------------------------------------------------------+
| NVIDIA-SMI 580.95.05              Driver Version: 580.95.05      CUDA Version: 13.0     |
+-----------------------------------------+------------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id          Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |           Memory-Usage | GPU-Util  Compute M. |
|                                         |                        |               MIG M. |
|=========================================+========================+======================|
|   0  NVIDIA GB10                    On  |   0000000F:01:00.0  On |                  N/A |
| N/A   40C    P8              4W /  N/A  | Not Supported          |      0%      Default |
|                                         |                        |                  N/A |
+-----------------------------------------+------------------------+----------------------+

+-----------------------------------------------------------------------------------------+
| Processes:                                                                              |
|  GPU   GI   CI              PID   Type   Process name                        GPU Memory |
|        ID   ID                                                               Usage      |
|=========================================================================================|
|    0   N/A  N/A            3254      G   /usr/lib/xorg/Xorg                      360MiB |
|    0   N/A  N/A            3426      G   /usr/bin/gnome-shell                    239MiB |
+-----------------------------------------------------------------------------------------+
```

- docker -v 查看 Docker 版本

```
leadtek@spark-b68e:~$ docker -v
Docker version 28.3.3, build 980b856
```

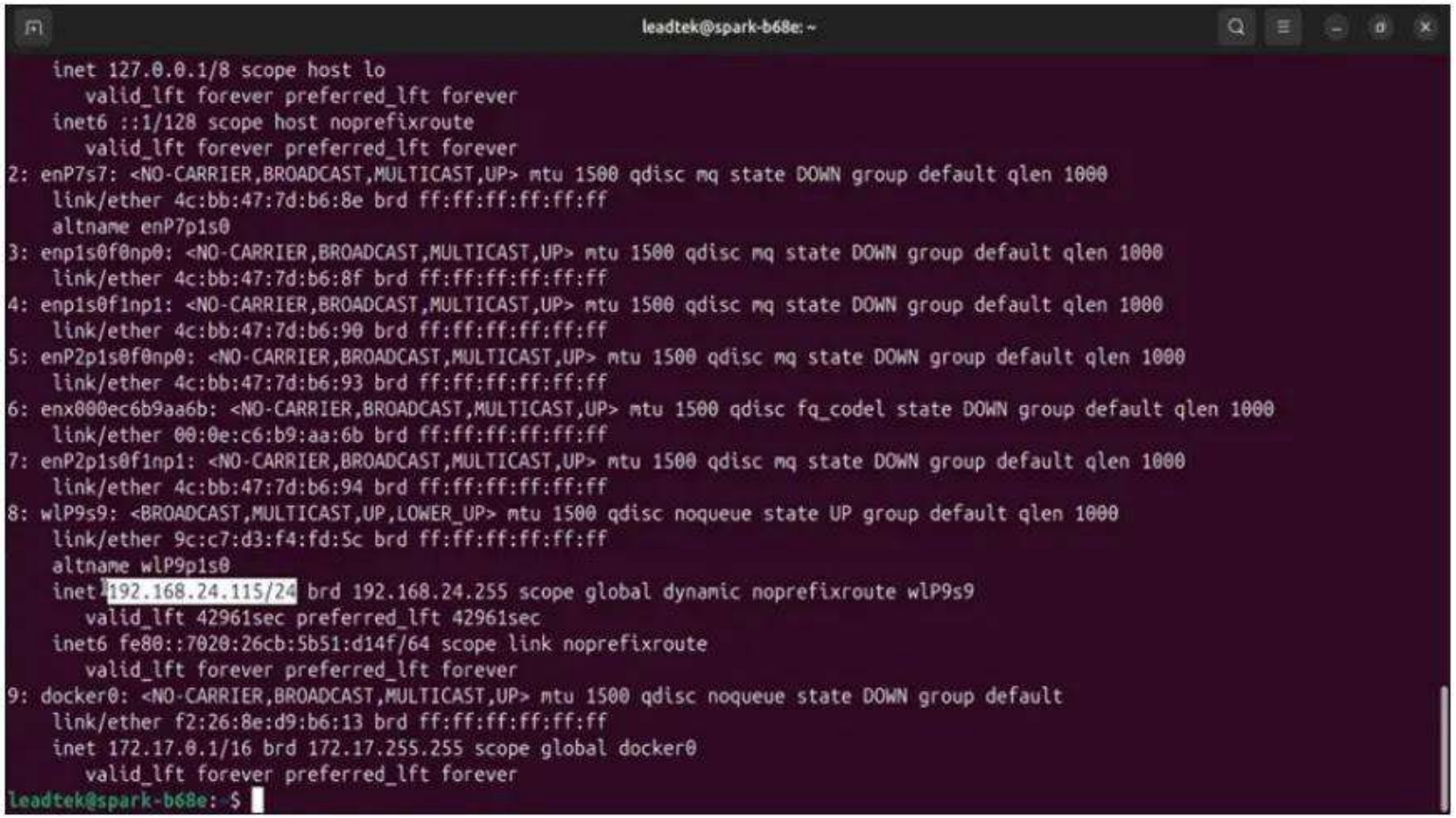
- nvcc -V 查看 CUDA 版本

```
leadtek@spark-b68e:~$ nvcc -V
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2025 NVIDIA Corporation
Built on Wed_Aug_20_01:57:39_PM_PDT_2025
Cuda compilation tools, release 13.0, V13.0.88
Build cuda_13.0.r13.0/compiler.36424714_0
```

五、 SSH指令列遠端登入

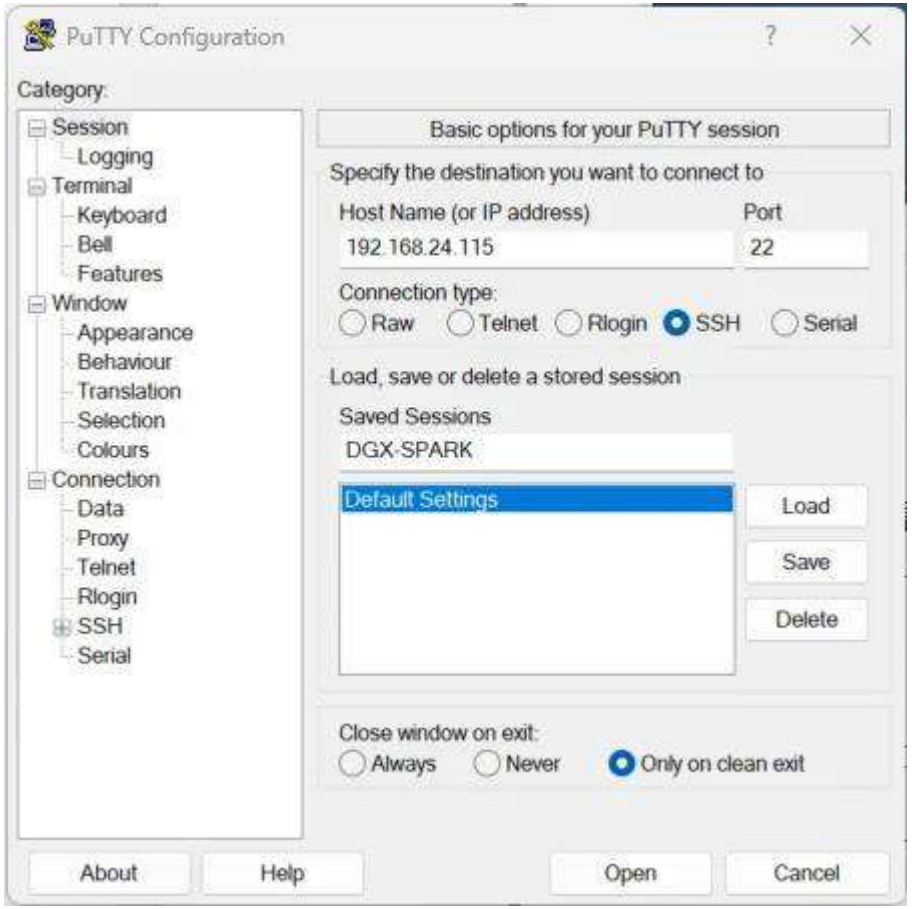
DGX Spark 預設可透過網路使用 SSH 從同網段其他電腦連線。

- 先輸入 ip a 查看本機 IP

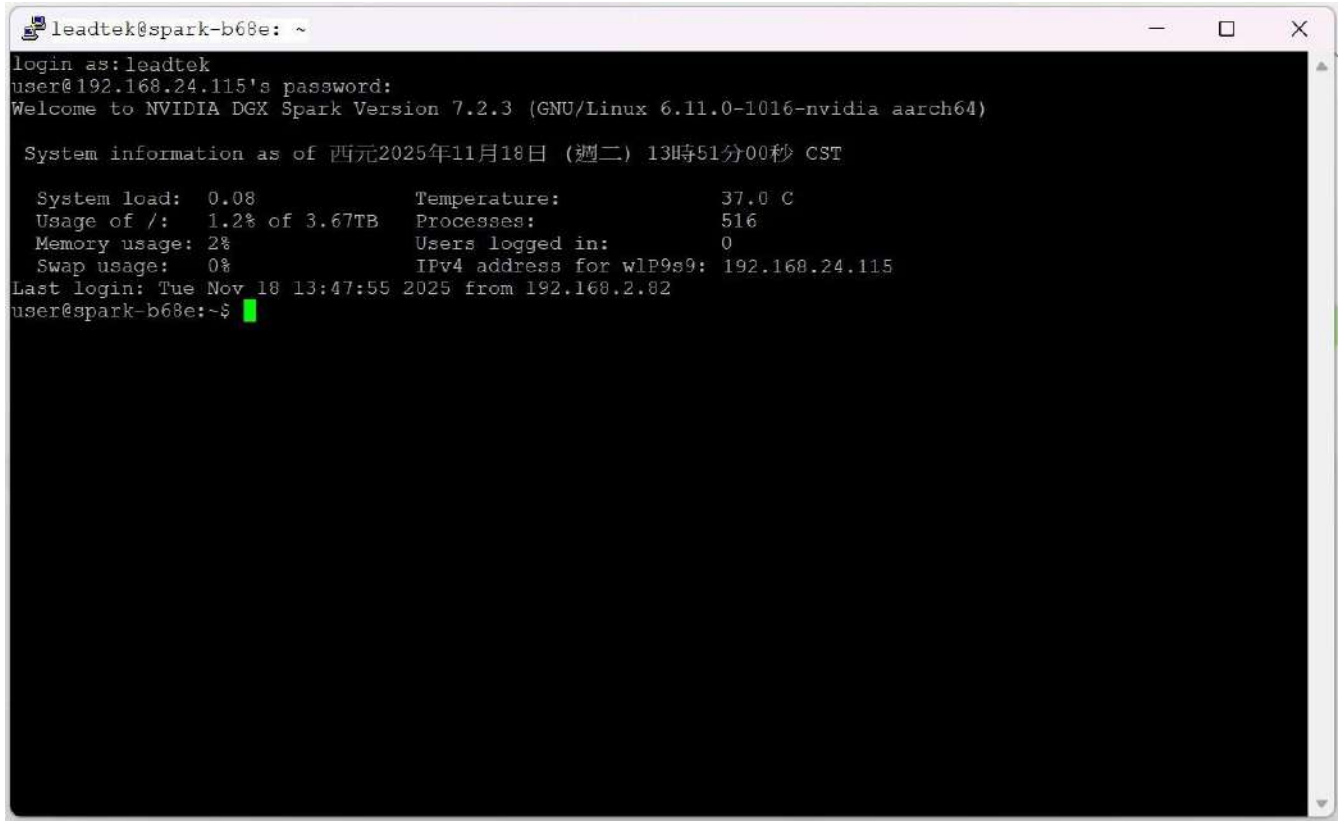


透過 SSH 工具遠端登入

- 以 PuTTY 為例，建立新連線，填入名稱、IP、埠號（預設 22），點擊Open後接著再輸入帳號及密碼。



- 點擊「**接受並儲存**」安全提示後即可使用遠端指令列操作 DGX Spark。



六、 Docker 設定

1.NVIDIA Container Runtime

NVIDIA Container Runtime 可讓 Docker 容器直接存取 GPU，以加速 AI / ML、CUDA 程式等 GPU 工作負載。
主要優點：

- 容器可無縫存取 GPU
- 自動管理驅動與函式庫
- 支援多 GPU
- 與主流容器編排平台相容

NVIDIA Container Runtime 與 NVIDIA Container Toolkit 相結合使用，後者提供了必要的元件，
以便為容器化應用動態配置並接入 GPU 設備和 CUDA 函式庫。
NVIDIA Container Toolkit 在 DGX Spark 系统中已預裝並完成配置，包括：

- NVIDIA Container Runtime
- Docker 整合
- GPU 裝置設定
- CUDA 函式庫配置

2.使用者群組設定

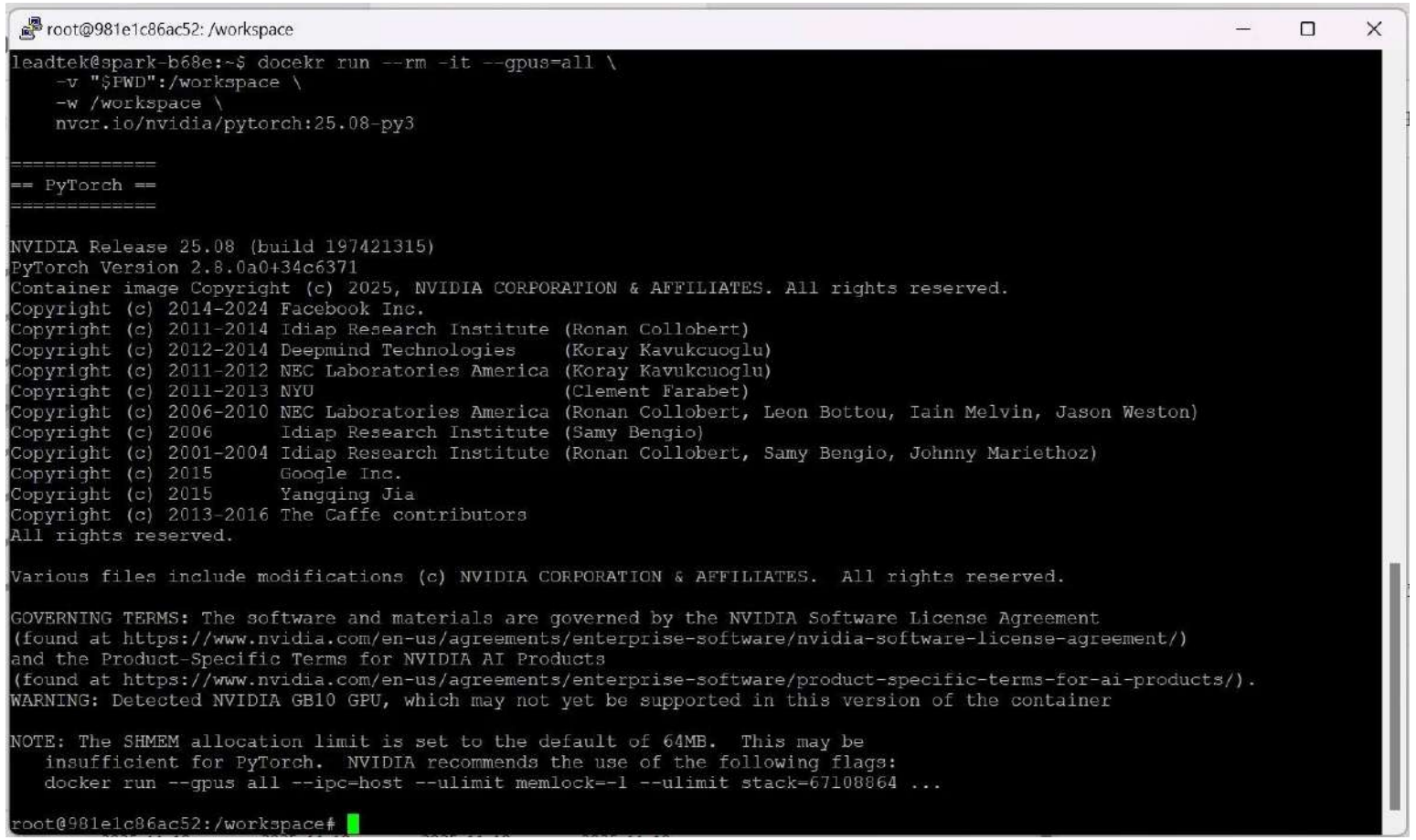
預設執行 Docker 需 sudo權限。將使用者加到 docker 群組，即可無需使用 sudo 直接執行 Docker 命令。

執行：
sudo usermod -aG docker $USER
newgrp docker

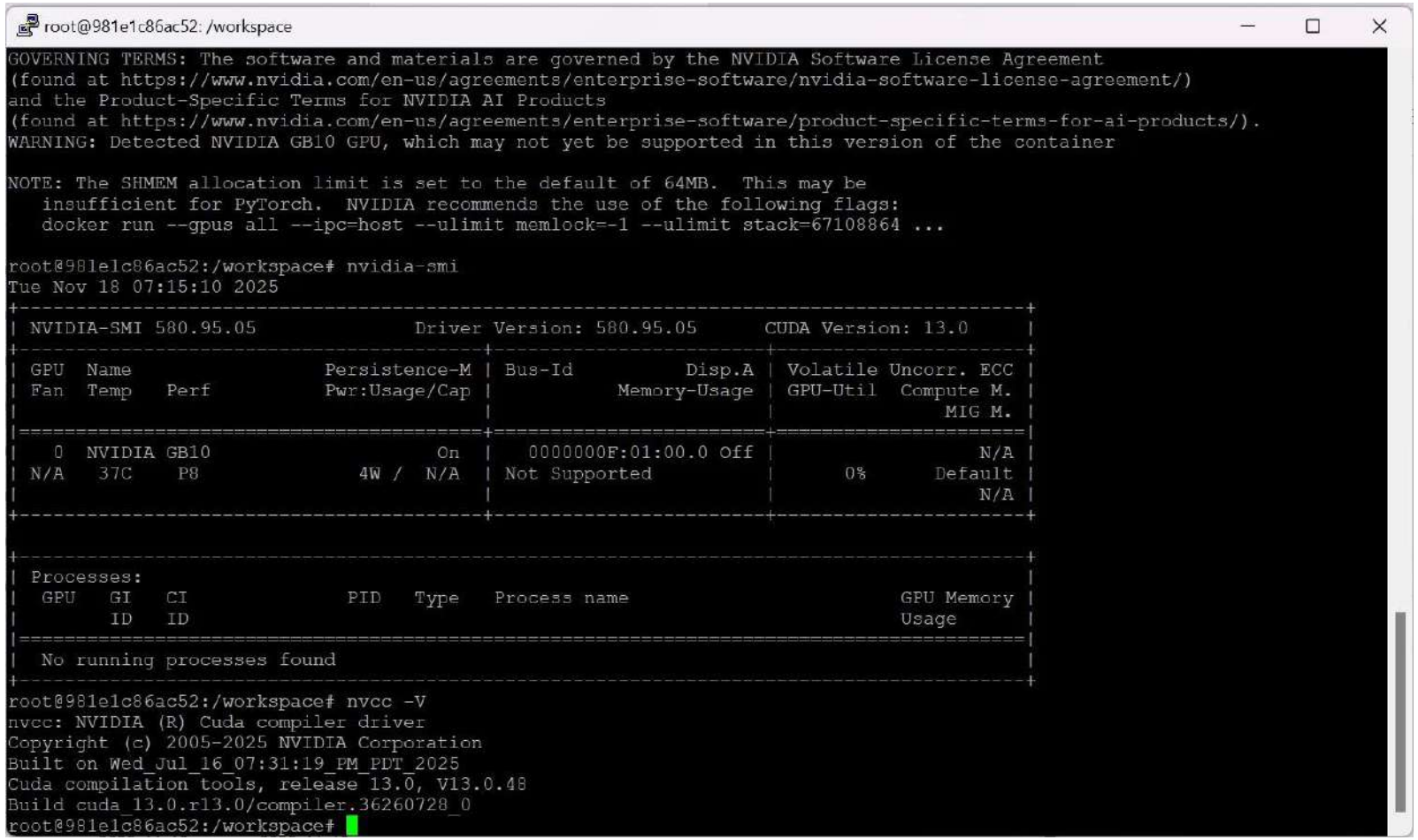
```
leadtek@spark-b68e:~$ sudo usermod -aG docker $USER
[sudo] password for leadtek:
leadtek@spark-b68e:~$ newgrp docker
leadtek@spark-b68e:~$
```

3.Docker GPU 功能測試

下載並執行 PyTorch 容器：
docker run --rm -it --gpus=all \
-v "$PWD":/workspace \
-w /workspace \
nvcr.io/nvidia/pytorch:25.08-py3



在容器內執行 nvidia-smi、nvcc -V 皆能正常輸出，即表示 GPU 與 CUDA 於 Docker 內可正常使用。



至此，您的 DGX Spark 已完成設定，可立即開始 AI 開發旅程！