

Nastavení GPU v Dockeru (GPU Passthrough)

Standardní Docker kontejnery nemají přímý přístup k hardwaru hostitele. Pro běh LLM/SLM modelů nebo trénování neuronových sítí je nutné zprovoznit tzv. **NVIDIA Container Toolkit** (dříve NVIDIA Docker).

1. Předpoklady

Před konfigurací Dockeru se ujistěte, že:

- Máte nainstalovány [aktuální ovladače GPU](#) na hostitelském stroji.
- Máte nainstalovaný **Docker Engine** (v26.x nebo novější).
- Na Linuxu máte nainstalovaný balíček curl a gpg.

2. Instalace NVIDIA Container Toolkit (Linux)

Toto je nejdůležitější krok, který přidá do Dockeru rozhraní nvidia runtime.

Přidání repozitáře

```
curl -fsSL https://nvidia.github.io/libnvidia-container/gpgkey | sudo gpg --dearmor -o /usr/share/keyrings/nvidia-container-toolkit-keyring.gpg \
&& curl -s -L https://nvidia.github.io/libnvidia-container/stable/deb/nvidia-container-toolkit.list | \
sed 's#deb https://#deb [signed-by=/usr/share/keyrings/nvidia-container-toolkit-keyring.gpg] https://#g' | \
sudo tee /etc/apt/sources.list.d/nvidia-container-toolkit.list
```

Instalace balíčku

```
sudo apt-get update
sudo apt-get install -y nvidia-container-toolkit
```

Konfigurace Dockeru

Musíte Dockeru říci, aby nový runtime používal:

```
sudo nvidia-ctl runtime configure --runtime=docker
sudo systemctl restart docker
```

3. Nastavení na Windows (WSL2)

V roce 2026 je situace na Windows jednodušší. Pokud používáte **Docker Desktop** a máte funkční **WSL2** (Windows Subsystem for Linux), podpora GPU je automaticky detekována.

- Stačí mít v nastavení Docker Desktop zaškrtnuto: „**Use the WSL 2 based engine**“.
- Není třeba instalovat Container Toolkit uvnitř WSL distribuít ručně, Docker Desktop to řeší za vás.

4. Ověření funkčnosti

Zkuste spustit testovací kontejner, který vypíše stav GPU:

```
docker run --rm --gpus all nvidia/cuda:12.4.0-base-ubuntu22.04 nvidia-smi
```

Pokud uvidíte tabulku se stavem grafické paměti, propojení funguje správně.

5. Použití v Docker Compose

Pro nasazení komplexnějších aplikací (např. [Open WebUI](#)) použijte v souboru `docker-compose.yml` sekci `deploy`:

```
services:
  ai-server:
    image: ollama/ollama
    deploy:
      resources:
        reservations:
          devices:
            - driver: nvidia
              count: all
              capabilities: [gpu]
```

Časté chyby (Troubleshooting)

Chyba	Řešení
„Unknown runtime: nvidia“	Nebyl spuštěn příkaz <code>nvidia-ctk runtime configure</code> .
„Could not select device with capabilities: [gpu],“	Ovladače na hostiteli jsou zastaralé nebo nekompatibilní s verzí Toolkitu.
Pomalý běh	Zkontrolujte, zda kontejner skutečně vidí GPU (použijte <code>nvidia-smi</code> uvnitř kontejneru).

— **Související dokumentace:**

- [Ovladače a CUDA](#)

- [Návod na Ollama v Dockeru](#)
- [GPU v Kubernetes \(K8s\)](#)

— **Správce:** @DevOps_Team **Poslední aktualizace:** 04. 01. 2026

From:

<https://serviceit.cz/> - **IT ENCYKLOPEDIE**

Permanent link:

https://serviceit.cz/doku.php?id=it:docker_gpu

Last update: **2026/01/04 16:43**

