

Laboratoris de la IA: on són i com els puc aprofitar per a la docència?

Els centres de càlcul de la UPC posen a disposició de l'alumnat i el professorat laboratoris de IA amb servidors d'alt rendiment equipats amb GPUs. Aquests recursos permeten accelerar el processament de dades i l'entrenament de models mitjançant biblioteques com CUDA, PyTorch i TensorFlow.

L'ús de GPUs en lloc de CPUs convencionals transforma radicalment l'experiència educativa en IA:

- **Acceleració de Càlcul Tensorial:** Les tasques que en una CPU trigarien hores (com la multiplicació de grans matrius en una xarxa neuronal), a la GPU es completen en minuts gràcies als CUDA Cores i Tensor Cores.
- **Inferència en Temps Real:** Permet provar models de detecció d'objectes en vídeo en temps real, una cosa impossible sense acceleració per hardware.
- **Recerca en IA Generativa:** Generació d'imatges (Stable Diffusion) o síntesi de veu, on la potència de càlcul defineix la qualitat i la velocitat del resultat.
- **Gestió de Memòria Compartida:** El fet de tenir múltiples GPUs en un mateix servidor permet als professors dissenyar pràctiques de Parallel Computing, on un sol model es reparteix entre diverses targetes.

Laboratoris d'IA a ETSETB

Els Serveis TIC de ETSETB ofereixen potència de càlcul distribuïda en diversos servidors especialitzats per a tasques intensives de GPU:

- **Servidor de Càlcul d'Alt Rendiment:**
 - 3 x NVIDIA GeForce RTX 3090 (24 GB)
 - 1 x NVIDIA RTX A5000 (16 GB)
 - 1 x NVIDIA Quadro RTX 5000 (16 GB)
- **Servidor de Gran Capacitat de Memòria de Vídeo:**
 - 2 x NVIDIA RTX A6000 (48 GB), ideal per a models de gran tamany.
- **Servidor d'Arquitectura Ada Lovelace:**
 - 1 x NVIDIA L4 (24 GB), optimitzada per a eficiència en inferència i IA.

Laboratoris d'IA a la FIB

Els Serveis TIC de la CNTIC ofereixen laboratoris d'IA per la docència de la FIB (graus, màsters, TFG i TFM) mitjançant diversos models de servei:

- **Clúster HPC "Boada"** (Gestió per cues SLURM):
 - **Configuració:** Nodes de còmput de CPU i un node especialitzat amb 8 GPUs.
 - **Ús:** Prioritat per a sessions de laboratori i pràctiques d'assignatures de la FIB. Els estudiants de TFG i TFM disposen d'accés en franges de baixa ocupació.
 - **Nota:** Properament s'ampliarà amb un node addicional de GPUs.
- **Servei GCED** (Virtualització VDI):
 - **Configuració:** Sistema basat en Ravada que permet l'ús de GPUs virtuals (vGPU) particionades.
 - **Ús:** Servei exclusiu per als laboratoris del Grau en Ciència i Enginyeria de Dades, garantint accés immediat i personalitzat sense cues.

- **Infraestructura de Projectes FIB** (En desplegament):
 - Nou servidor amb 4 GPUs destinat a l'explotació de serveis i projectes docents de la facultat.
- **Recursos Presencials i Préstec:**
 - **Aules Especialitzades:** Dues aules informàtiques d'altres prestacions preparades per a IA i Alta Computació (A5S105 i C6S308).
 - **Portàtils de Préstec:** Lot de 27 portàtils equipats amb GPUs de consum per a ús autònom de l'estudiantat.

Com puc aprofitar aquests laboratoris per la docència?

1. Entrenament de Models de Deep Learning

Els estudiants poden treballar amb conjunts de dades voluminosos en àrees com:

- **Computer Vision:** Entrenament de xarxes neuronals convolucionals (CNN) per al reconeixement d'imatges o la segmentació d'imatges mèdiques.
- **Natural Language Processing (NLP):** Processament de grans corpus de text per a anàlisi de sentiment o traducció.

Permetent realitzar pràctiques en aprenentatge profund on s'entrenen xarxes neuronals convolucionals (CNN) per a reconeixement d'imatges o segmentació mèdica.

2. Fine-tuning de Models de Llenguatge (LLMs)

Els estudiants poden realitzar pràctiques de **Fine-tuning** (o ajust fi) que és el procés de realitzar un entrenament addicional a un model de llenguatge que ja ha estat prèviament entrenat (un Pre-trained Model), utilitzant un conjunt de dades més petit i específic.

3. Computació Distribuïda i Gestió de Cues (Clúster Boada)

L'existència d'un clúster HPC amb SLURM permet als estudiants anar més enllà de l'entrenament d'un model simple i aprendre la gestió de recursos a escala industrial.

- **Optimització d'Hiperparàmetres en Paral·lel:** Els estudiants de IA o de la FIB poden llançar experiments simultanis en diferents nodes o GPUs del Boada.
- **Gestió de Cues i Prioritats:** Aprenentatge pràctic sobre com programar tasques d'alt còmput sense col·lapsar el sistema, entenent conceptes com el fair share i la gestió de treballs en batch.

4. Virtualització de GPU per a Ciència de Dades (Servei GCED)

L'ús de vGPUs (GPU virtuals) i Ravada en el grau d'Enginyeria de Dades obre escenaris de treball immediat i personalitzat.

- **Entorns de Desenvolupament Interactius (VDI):** Gràcies a la virtualització, cada estudiant té el seu propi escriptori remot amb una fracció de memòria de GPU garantida. Això és ideal per a sessions de laboratori on es requereix visualització de dades complexa o anàlisi exploratòria immediata sense esperar cues de processament.

5. Desplegament de Serveis de IA en Producció (Projectes FIB)

Aquest nou recurs permet tancar el cercle del cicle de vida de la IA (MLOps), passant del laboratori a l'explotació real.

6. Computació de Proximitat i "Edge AI" (Aules i Portàtils de Préstec)

L'accés a hardware físic directe (GPUs de consum) permet explorar aplicacions on la latència i el control del hardware són crítics.

- **Desenvolupament de Videojocs i Renderització:** Ús de les GPUs de les aules per a projectes que requereixen Ray Tracing en temps real o l'ús de motors com Unreal Engine o Unity, on la GPU no només entrena IA, sinó que genera entorns visuals complexos.
- **Protoclatge de "Edge Computing" amb Portàtils:** Els portàtils de préstec permeten als estudiants sortir del laboratori i recollir dades en el món real (per exemple, visió per computador amb la càmera del portàtil en moviment) i processar-les localment sense dependre d'una connexió al clúster, facilitant TFGs més experimentals i de camp.