GPU-Compatible Library Dependencies (PowerPC / H200)

This document lists the key deep learning libraries (PyTorch, Transformers, DeepSpeed) and their associated dependencies that must be addressed for compatibility on PowerPC with CUDA 12.2 and H200 GPUs.

# Dependency Matrix by Library

|  |  |  |
| --- | --- | --- |
| Primary Library | Key Dependencies | Notes |
| PyTorch | cudatoolkit, cudnn, nccl, torchvision, torchaudio, torchtext, pytorch-cuda | Make sure CUDA version (e.g., 12.2) aligns with PyTorch build. NCCL is important for multi-GPU setups. |
|  | nvidia-ml-py3, apex (optional) | apex used for mixed precision. Needs to be built from source on PowerPC. |
| Transformers | torch, tokenizers, accelerate, datasets, huggingface\_hub, peft | tokenizers is Rust-based and may need special handling on PowerPC. accelerate is critical for training. |
|  | safetensors, evaluate, sentencepiece, scipy, numpy, protobuf | Required for model loading, evaluation, and export. |
| DeepSpeed | torch, ninja, transformers (optional), mpi4py, deepspeed dependencies | Needs CUDA dev tools to compile C++ extensions. NCCL and ninja required for building kernels. |
|  | aiohttp, psutil, py-cpuinfo, tqdm, packaging | Runtime dependencies. |

Installation Order Recommendation:  
1. Set up CUDA + PyTorch first and verify GPU functionality.  
2. Install Transformers and Hugging Face ecosystem next.  
3. Compile or install DeepSpeed last, as it hooks into PyTorch and requires compilation.