NVIDIA L4 - Llama-3.1-8B-Instruct