Homework for session 7: Sukruth Somappa

Theory homework

1. **What are the key architectural features that make these systems suitable for AI workloads?**
   Large memory on chip, specialized hardware for tensor manipulations, parallel processing capability

2. **Identify the primary differences between these AI accelerator systems in terms of their architecture and programming models**

   Cerebras: Parallel workflow with processing elements with own memory
   Graphcore: Interconnected processing tiles needing two phase operation – computation and communication using Bulk Synchronous Parallelism
   Groq's: Deterministic execution
   Sambanova: Multi-tiered memory architecture efficient for large data

3. **Based on hands-on sessions, describe a typical workflow for refactoring an AI model to run on one of ALCF's AI testbeds (e.g., SambaNova or Cerebras). What tools or software stacks are typically used in this process?**

   Preparing data, Model creation – training, inference, Training
   Using Pytorch, Tensorflow, etc.

4. **Give an example of a project that would benefit from AI accelerators and why?**
   For data mining: Sometimes it can be hard to extract relevant data buried in bureaucratic documents, for example in real estate, health care, etc. AI accelerators can be used to extract data parallelly and in quick time.