DeepSpeed Hands-on session at KAUST

Ammar Ahmad Awan (Microsoft)

Mohsin A. Shaikh (KAUST)

- Code examples for today are here:

https://github.com/kaust-rccl/deepspeed_workshop.git

- Contains jobscripts and patch file

- For Megatron-Deepspeed, copy the Jobscripts to Megatron-DeepSpeed
    - https://github.com/microsoft/Megatron-DeepSpeed.git

Session 1: Introduction to DeepSpeed

30 mins – HelloDeepSpeed

Fork the repository:https://github.com/microsoft/DeepSpeedExamples

1. Go to HelloDeepSpeed folder.
2. Run the train_bert_ds.py
3. Edit the file and play with
    1. ZeRO stages
    2. ZeRO offload_optim and offload_param
    3. Optional – Add gradient/activation checkpointing

Session 2: Megatron-DeepSpeed

Clone/fork the repository: https://github.com/microsoft/megatron-deepspeed

- Go to examples/azure folder
- Run the benchmark model script
  - Investigate certain options and change them
    - Batch size per GPU
    - ZeRO stages: 0, 1, 2, 3 – validate your understanding from the intro. slides
    - ZeRO offload and CPUAdam
    - Activation checkpointing

Session 3: DeepSpeed Inference

- HuggingFace exercise (30 mins)
  - Text generation examples

- Stable Diffusion (30 mins)
  - Image generation examples