Frederick National Laboratory
for Cancer Research

NIH CANDLE Workshop
Running CANDLE Benchmarks on Biowulf

George Zaki

HPC group, DSITP

April 19, 2017

DEPARTMENT OF HEALTH AND HUMAN SERVICES • National Institutes of Health • National Cancer Institute

Frederick National Laboratory is a federally funded research and development center operated by Leidos Biomedical Research, Inc., for the National Cancer Institute.

What this session is about?

- I would like to run a baseline CANDLE benchmark in batch mode on Biowulf.
- I would like to integrate a benchmark with my pipeline.
- I would like to edit a benchmark and do a quick evaluation in an interactive session on Biowulf.

Student Account Setup

- Complete instructions are on github: https://github.com/ECP-CANDLE/Workshop

- If you want to use the student account today to train p1b1: replace `student1` with your account. Use: P@ssw0rd:

```
ssh student1@helix.nih.gov

ssh student1@biowulf.nih.gov

/data/classes/candle/setup.sh

sbatch --partition=student --mem=50g --cpus-per-
task=8/data/classes/candle/jobrequest.sh
```

- To run a different benchmark, edit:

```
~/candle-jobscript.sh
```

Biowulf User Setup

- You can follow the same student setup instructions!
  Just omit the student partition.
  This will create:
  - `/data/`whoami`/candle/`
  - `~/candle-jobscript.sh`

- If you would like to edit the benchmarks, run in interactive mode, submit your training job to a specific node type (K80).
  Then follow up the next few slides

Candle Benchmarks Requirements

- Software requirements
  - Keras
  - Other supporting python packages (hd5, opencv, etc)

- Hardware requirement
  - 50GB of memory (this number might change in the future)
  - CPU or CPU nodes (All benchmarks were tested on CPU nodes). Some benchmarks requires more memory than what is available in K20 GPU nodes.
    - Might need to wait for K80 nodes
  - 30mins to 30 hours (check every benchmark's README)

Software Setup 1

- All the required python packages are installed in a singularity container image.
  - /data/classes/candle/keras.img

- Clone the benchmarks to a local directory:
  - `export CANDLE_BENCH=/data/`whoami`/candle-benchmakrs`
  - `git clone https://github.com/ECP-Candle/benchmarks $CANDLE_BENCH`

Software Setup 2

- To use the singularity image, first load the singularity module:
  - `sinteractive –mem=50g --cpus-per-task=8`
  - `module load singularity`

- In interactive mode run:
  - `singularity **shell** --bind $CANDLE_PATH:/Benchmarks /data/classes/candle/keras.img`
  - `python /Benchmarks/Pilot1/P1B1/p1b1_baseline_keras2.py`

Software Setup 3

- In batch script, you would like to execute multiple commands within your singularity container. To do that run:
  - `singularity` **`exec`** `--bind $CANDLE_PATH:/Benchmarks /data/classes/candle/keras.img /path/to/`**`candle-jobscript.sh`**

- While candle-jobscript.sh contains the commands you would like to execute. For example,
  - `python /Benchmarks/Pilot1/P1B1/p1b1_baseline_keras2.py`

  - Here the `/Benchmarks` directory is only seen within your container.

Hardware Setup

- Hardware request can be done with your sbatch or sinteractive:
  - --mem=50g #To request memory cpu
  - --gres=gpu:k20x:1 # To request a GPU node
  - --time=24:00:00 # To request a specific time for the job
  - --cpus-per-task=8 # To ask for a specific number of cpus

- For more information about job submission on Biowulf: https://hpc.nih.gov/docs/userguide.html

Candle Benchmarks Profiles

Benchmark	Time	Memory
P1B1	10 mins	10 GB
P1B2	10 mins	4 GB
P1B3	30 hours	5 GB
P2B1	15 mins	2 GB
P2B2	10 mins	2 GB
P3B1	10 mins	1 GB
P3B2	5 hours	7 GB

Disclaimer: These numbers are specific to the particular system used. In this case, the process was run on the CPU (8 cores) only and did not use GPU Resources.

Intel® Xeon® CPU E5-2643 @ 3.3GHz

Questions

- Contact info:
  - george.zaki@nih.gov