DS 7347
High-Performance Computing (HPC) and Data Science
Session 18

Robert Kalescky
Adjunct Professor of Data Science
HPC Research Scientist

June 23, 2022

Research and Data Sciences Services
Office of Information Technology
Center for Research Computing
Southern Methodist University

Outline

Session Question

Session Question

In what ways are benchmarks useful?

Benchmarking

Performance Benchmarking

- Understand application or worflow performance
- Provides understanding of how performance changes as development continues or different hardware is used
- The idea is very simple, time the portions of the workflow that are useful to be benchmarked
  - In lower-level code you can insert timing statements or use one of many libraries available
  - In higher-level scripts you can simply use the `time` command

Common Types of Benchmarking

- Entire workflow
- Major subtasks of your workflow
- Performance critical tasks:
  - Compute intensive tasks
  - IO intensive tasks

Best Practices

- Repeat the benchmarking multiple times looking at the average and variance
- Isolate the area of interest from the normal fluctuations of a shared compute resource
- Document specific details of the job's software stack, the resources being used, and the timings

Group Benchmarking Example

- GEMM on different node types and GPUs
- See `blas_benchmark` directory.

Readings and Assignments

Project

- Prototype of multi-job Slurm submit script.
- Commit the script to your project repo
- Due 12:00 AM Central, Tuesday, June 21, 2022

Readings and Assignments

Readings

None

Project

- Implement one subtask of your workflow using “easiest” installation path.
- Benchmark the tasks.
- Commit the details to your project repo
- Due 12:00 AM Central, Tuesday, June 28, 2022