Benchmarking

APRIL 30TH, 2020

What is Benchmarking?

- Benchmarking is running a program to assess the relative performance of a something
  - Measures the whole time for the program
  - Compares different hardware/architecture
  - Can verify correctness of software

- Profiling is running tests to assess the relative performance of individual pieces of a program

Why Benchmark?

- Small scale performance to show large scale performance

- General performance

- Proteus-specific performance

  - AMD vs INTEL

  - https://proteusmaster.urcf.drexel.edu/urcfwiki/index.php/Proteus_Hardware_and_Software

- Test correctness

Things to Consider

- Number of threads

- Number of nodes

- Varying slot request
  - Can greatly vary time to complete

- Hybrid MPI

- Hardware

- External code
  - May not be optimized

Know What Your Code Does

- Memory?

- Storage Space?

- Run Time?

- Sub-processes?

Understanding Your Code

- qstat –j xxxxxx

- qacct –j xxxxxx

- Ganglia https://proteusmaster.urcf.drexel.edu/ganglia-proteus/

- Documentation

```
(benchmark) [cwf25@proteusi01 benchmarking]$ qacct -j 534253
================================================================
qname           all.q
hostname        ic05n01.cm.cluster
group           urcfcoopGrp
owner           cwf25
project         urcfcoopPrj
department      defaultdepartment
jobname         test.sh
jobnumber       534253
taskid          undefined
account         sge
priority        0
cwd             /mnt/HA/groups/urcfcoopGrp/opt/workshops/benchmarking
submit_host     proteusi01.cm.cluster
submit_cmd      qsub test.sh
qsub_time       04/28/2020 14:51:24.447
start_time      04/28/2020 14:51:26.114
end_time        04/28/2020 14:52:13.230
granted_pe      shm
slots           2
failed          0
deleted_by      NONE
exit_status     0
ru_wallclock 47.116
ru_utime        52.761
ru_stime        1.650
ru_maxrss       216772
ru_ixrss        0
ru_ismrss       0
ru_idrss        0
ru_isrss        0
ru_minflt       310744
ru_majflt       380
ru_nswap        0
ru_inblock      326688
ru_oublock      176
ru_msgsnd       0
ru_msgrcv       0
ru_nsignals 0
ru_nvcsw        37586
ru_nivcsw       11801
wallclock       48.127
cpu             54.411
mem             27.967
io              0.291
iow             10.560
ioops           42174
maxvmem         2.556G
maxrss          542.598M
maxpss          474.664M
arid            undefined
jc_name         NONE
```

qstat and qacct – Important Fields

- hostname – node where the task ran

- end_time – date-time when task ended

- wallclock – wallclock time that job spent in running state

- io – amount of data transferred in GiB

- iow – io wait time in seconds

- maxvmem – maximum amount of memory used

- maxrss – maximum amount of memory used by an individual process

```
(benchmark) [cwf25@proteusi01 benchmarking]$ qacct -j 534253
==============================================================================
qname          all.q
hostname       ic05n01.cm.cluster
group          urcfcoopGrp
owner          cwf25
project        urcfcoopPrj
department     defaultdepartment
jobname        test.sh
jobnumber      534253
taskid         undefined
account        sge
priority       0
cwd            /mnt/HA/groups/urcfcoopGrp/opt/workshops/benchmarking
submit_host    proteusi01.cm.cluster
submit_cmd     qsub test.sh
qsub_time      04/28/2020 14:51:24.447
start_time     04/28/2020 14:51:26.114
end_time       04/28/2020 14:52:13.230
granted_pe     shm
slots          2
failed         0
deleted_by     NONE
exit_status    0
ru_wallclock   47.116
ru_utime       52.761
ru_stime       1.650
ru_maxrss      216772
ru_ixrss       0
ru_ismrss      0
ru_idrss       0
ru_isrss       0
ru_minflt      310744
ru_majflt      380
ru_nswap       0
ru_inblock     326688
ru_oublock     176
ru_msgsnd      0
ru_msgrcv      0
ru_nsignals    0
ru_nvcsw       37586
ru_nivcsw      11801
wallclock      48.127
cpu            54.411
mem            27.967
io             0.291
iow            10.560
ioops          42174
maxvmem        2.556G
maxrss         542.598M
maxpss         474.664M
arid           undefined
jc_name        NONE
```

Two Bio Programs

BLAST+

- BLAST+ is a command line program that finds regions of similarity between genetic sequences
- Current available implementation on Proteus is 2.6
  - https://proteusmaster.urcf.drexel.edu/urcfwiki/index.php/Compiling_NCBI_BLAST

MERCAT

- MerCat is a python program for k-mer counting.
  - k-mer counting means counting the number of times each pattern of k length occurs
- Available at https://github.com/pnnl/mercat.git

Let's test them!