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

•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?

		[cwf25@proteusi01 benchmarking]\$ qacct −j 534253
	qname hostname	all.q ic05n01.cm.cluster
	group owner	urcfcoopGrp cwf25
	project department	urcfcoopPrj defaultdepartment
	jobname	test.sh
Undarctanding Vour Codo	jobnumber taskid	534253 undefined
Understanding Your Code	account priority	sge θ
	cwd submit_host	/mnt/HA/groups/urcfcoopGrp/opt/workshops/benchmarking proteusi01.cm.cluster
	submit_cmd qsub_time	qsub test.sh 04/28/2020 14:51:24.447
	<pre>start_time end_time</pre>	04/28/2020 14:51:26.114 04/28/2020 14:52:13.230
•qstat—j xxxxxx	granted_pe slots	shm 2
	failed deleted by	e NONE
	exit_status ru wallclock	
•qacct –j xxxxxx	ru_utime	52.761 1.650
	ru_stime ru_maxrss	216772
	ru_ixrss ru_ismrss	
Canglia https://protousmaster.uref.drovel.edu/ganglia.protous/	ru_idrss ru_isrss	
 Ganglia <u>https://proteusmaster.urcf.drexel.edu/ganglia-proteus/</u> 	ru_minflt ru_majflt	310744 380
	ru_nswap ru_inblock	0 326688
	ru_oublock ru_msgsnd	176 0
 Documentation 	<pre>ru_msgrcv ru_nsignals</pre>	
	ru_nvcsw ru nivcsw	37586 11801
	wallclock cpu	48.127 54.411
	mem io	27.967 0.291
	iow	10.560
	ioops maxvmem	42174 2.5566
	maxrss maxpss	542.598M 474.664M
	arid jc_name	undefined 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 534253
	jobnumber taskid	undefined
	account	sge
	priority	9 9
	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	shim
	slots	2
	failed	θ
	deleted_by	NONE
	exit_status	
	ru_wallclock	47.116
	ru_utime	52.761
	ru_stime	1.650
	ru_maxrss	216772
	ru_ixrss	
	ru_ismrss	
~	ru_idrss	
S	ru_isrss	
	ru_minflt	310744
	ru_majflt	380
	ru_nswap	0
	ru_inblock	326688
	ru_oublock	176
	ru_msgsnd	0
	ru_msgrcv	0 0
	ru_nsignals	
	ru_nvcsw ru nivcsw	37586 11801
	wallclock	48.127
	сри	54.411
	mem	27.967
	io	0.291
	iow	10.560
	ioops	42174
	maxvmem	2.5566
	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
 - <u>https://proteusmaster.urcf.drexel.edu/urcfwiki/i</u> ndex.php/Compiling NCBI BLAST

MERCAT

- •MerCat is a python program for kmer counting.
 - k-mer counting means counting the number of times each pattern of k length occurs
- •Available at <u>https://github.com/pnnl/mercat.gi</u> <u>t</u>

Let's test them!