

GPGPUs

	P100 on E5-2640	P100 on Silver-4110	V100
Hardware			
Host	max-wng005	max-p3ag028	
CPU	E5-2640 v4 @ 2.40GHz cpu MHz : 1202.343 cache size : 25600 KB	Silver 4114 CPU @ 2.20GHz cpu MHz : 800.000 cache size : 14080 KB	
Memory	256GB	768GB	
GPU	P100-PCIE-16GB	P100-PCIE-16GB	
BUS & Numa node	bus:0x02 numa: 2	bus:0x86 numa: 1	
CUDA Tests			
Bandwidth (MB/s)	Host Device 11709	12049	
	DeviceHost 12849	12863	
	DeviceDevice 500636	500300	
p2pBandwidthLatencyTest (GB/s)	UNI P2P Disabled 346	504	
	UNI P2P Enabled 347	504	
	BI P2P Disabled 358	512	
	BI P2P Enabled 357	513	
convolutionFFT2D (Mpix/s)	built-in R2C / C2R 6088	6088	
	custom R2C / C2R 6144	6107	
	updated custom R2C / C2R 6069	7812	
simpleMultiCopy (GB/s)	Host Device 11.9	12.2	
	DeviceHost 12.7	12.7	
	Kernel 417.1	1248.0	
	Serialized exec 10.6	12.0	
	4 Streams 19.4	20.6	
matrixMul (GFlop/s)	420.0	1766	
matrixMulCUBLAS (GFlop/s)	2321.7	5579	
Benchmarks			
shmembench	using 32bit operations : 1777.72 GB/sec	7614.55 GB/sec	
	using 64bit operations : 1815.44 GB/sec	7890.97 GB/sec	
	using 128bit operations : 1816.02 GB/sec	7662.91 GB/sec	
constbench	using 32bit operations : 650.68 GB/sec	2778.73 GB/sec	
	using 64bit operations : 2247.59 GB/sec	9522.36 GB/sec	
	using 128bit operations : 2402.28 GB/sec	10218.13 GB/sec	

cachebench	Read only accesses:		
	int1: 1000.92 GB/sec	int1: 2121.66 GB/sec	
	int2: 2106.63 GB/sec	int2: 2310.82 GB/sec	
	int4: 2379.41 GB/sec	int4: 2379.64 GB/sec	
	max: 2379.41 GB/sec	max: 2379.64 GB/sec	
	Read-write accesses:		
int1: 2229.74 GB/sec	int1: 2230.04 GB/sec		
int2: 2210.64 GB/sec	int2: 2211.12 GB/sec		
int4: 2127.50 GB/sec	int4: 2126.41 GB/sec		
max: 2229.74 GB/sec	max: 2230.04 GB/sec		
gpu-burn (Gflops/s)	7990	7990	
Others			
Tensorflow Resnet 50	images/sec: 214.9	images/sec: 210.41	