17.8-Gbps Disk-to-Disk File Copies Achieved Via Workstations Costing Less Than $9,000

- As part of plans to assess the throughput performance of wide-area file transfer applications, GSFC's High End Computer Network Team specified and assembled workstations that individually costs less than $9,000 and are capable of over 17.8 gigabits per second (Gbps) disk-to-disk file copying.
- Each workstation consists of a 3.2-GHz single-processor (quad core) Intel Core i7 (Nehalem) with four HighPoint RocketRaid 4320 RAID disk controllers and a Myricom 2-port 10 Gigabit Ethernet network interface card in the PCIe Gen2 slots of a Asus P6T6 WS Revolution motherboard. Each RAID controller hosts eight Western Digital WD5001AALS 500-gigabyte disks.
- Over 17.8-Gbps disk-to-disk file-copying throughput between two of the workstations was measured using the nuttscp (www.nuttcp.net) file copying tool.
- While SSD technology is next to be investigated, parallelism of multiple cores and multiple streams is likely to be key to going to 40-Gbps and beyond, since individual cores are not getting significantly faster.

***Figure*:** Right case houses Core i7 cores, DDR3 memory, NIC, two "internal" controllers each with eight disks and two "external" controllers; left case houses sixteen SAS-connected disks.

POC: Bill Fink, William.E.Fink@nasa.gov, (301) 286-7924, GSFC Computational and Information Sciences and Technology Office

Introduction To GSFC High End Computing 20, 40 &100 Gbps Network Testbeds

Precursor Tests of "C-Systems" (to show the individual components have the necessary muscle) [Source: Bill Fink/GSFC]

- Disk I/O speeds via dd reads (of=/dev/null) & writes (if=/dev/zero)
  - Read: 68719476736 bytes (69 GB) copied, 25.8791 s, 2.7 GB/s
  - Write: 68719476736 bytes (69 GB) copied, 26.8676 s, 2.6 GB/s

- 2x10-GigE via "nuttcp application bonding"
  - TX: 19805.8537 Mbps 34 %TX 59 %RX 0 retrans 0.11 msRTT
  - RX: 19808.7300 Mbps 39 %TX 53 %RX 0 retrans 0.11 msRTT

Introduction To GSFC High End Computing 20, 40 &100 Gbps Network Testbeds

Nuttscp Sample Test Results Between Two "C-Systems" (1-of-7) [Source: Bill Fink/GSFC]

- One 64-GB file copy (between four RAID5 disk controllers nested as RAID50 hosted on each C-system in a LAN testbed)
  - Configuration settings:
    - LRO enabled
    - eth2,3 interrupts on CPU0
    - nuttcp application running on CPU1
    - 4xHPT RAID5 interrupts running on CPU2
    - md RAID50 across above
  - Get: 10273.4125 Mbps 52 %TX 99 %RX 0 retrans 0.11 msRTT
  - Put: 10311.2700 Mbps 52 %TX 99 %RX 0 retrans 0.11 msRTT
- Houston, we have a problem! We're definitely not firing on all cylinders. It's obvious what the problem is, namely that the receiver CPU is totally saturated. To go faster is going to require nuttcp using multiple cores in parallel….

Introduction To GSFC High End Computing 20, 40 &100 Gbps Network Testbeds

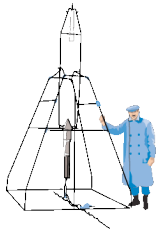
Nuttscp Sample Test Results Between Two "C-Systems" (2-of-7) [Source: Bill Fink/GSFC]

- One 64-GB file copy similar to "1-of-7" but only one side's RAID50 is real
  - Configuration settings: same as in "1-of-7"
  - Get from RAID50 to /dev/null:
    17324.4416 Mbps 98 %TX 49 %RX 0 retrans 0.11 msRTT
  - Put from /dev/zero to RAID50:
    10129.7218 Mbps 27 %TX 99 %RX 0 retrans 0.11 msRTT
- So, the immediate 20-Gbps challenge is primarily on the write side….

Introduction To GSFC High End Computing 20, 40 &100 Gbps Network Testbeds

Nuttscp Sample Test Results Between Two "C-Systems" (3-of-7) [Source: Bill Fink/GSFC]

- Two 64-GB file copy (between four RAID5 disk controllers nested as RAID50 hosted on each C-system in a LAN testbed)
  - Configuration settings: same as in "1-of-7" **plus**
    - nuttcp application running on CPU3
  - Put file1:
    7184.8745 Mbps 41 %TX 71 %RX 0 retrans 0.11 msRTT
  - Put file2:
    7082.7940 Mbps 46 %TX 70 %RX 0 retrans 0.11 msRTT
  - Aggregate throughput:
    14267.6685 Mbps
- Better; but there was a lot of disk head contention seeking back and forth between the two files



Introduction To GSFC High End Computing 20, 40 &100 Gbps Network Testbeds

Nuttscp Sample Test Results Between Two "C-Systems" (4-of-7) [Source: Bill Fink/GSFC]

- A slight variation of "3-of-7", using individual 10-GigE nuttcp streams across individual 10-GigEpaths
  - Put file1:
    7136.7905 Mbps 39 %TX 72 %RX 0 retrans 0.11 msRTT
  - Put file2:
    7123.8836 Mbps 39 %TX 72 %RX 0 retrans 0.11 msRTT
  - Aggregate throughput:
    14260.6741 Mbps
- Basically the same result as "3-of-4"

Introduction To GSFC High End Computing 20, 40 &100 Gbps Network Testbeds

Nuttscp Sample Test Results Between Two "C-Systems" (5-of-7) [Source: Bill Fink/GSFC]

- Splitting the one RAID50 into two separate RAID50s to avoid the disk head seeking contention
  - Configuration settings:
    - LRO enabled
    - eth2,3 interrupts on CPU0
    - nuttcp s2 application running on CPU1
    - 2xHPT RAID5 interrupts running on CPU2
    - first md RAID50 across above
    - 2xHPT RAID5 interrupts running on CPU2
    - second md RAID50 across above
    - nuttcp s1 application running on CPU3
  - Put file1/s1:
    9318.3251 Mbps 55 %TX 92 %RX 0 retrans 0.11 msRTT
  - Put file2/s2:
    7960.6777 Mbps 47 %TX 79 %RX 0 retrans 0.10 msRTT
  - Aggregate throughput:
    17279.0028 Mbps

Introduction To GSFC High End Computing 20, 40 &100 Gbps Network Testbeds

Nuttscp Sample Test Results Between Two "C-Systems" (6-of-7) [Source: Bill Fink/GSFC]

- Similar to "5-of-7" but moving the last 2 HPT RAID5 interrupts to CPU 0, so stream s2could have the same advantage as stream s1
  - Configuration settings:
    - LRO enabled
    - eth2,3 interrupts on CPU0
    - 2xHPT RAID5 interrupts running on CPU0
    - second md RAID50 across above
    - nuttcp s2 application running on CPU1
    - 2xHPT RAID5 interrupts running on CPU2
    - first md RAID50 across above
    - nuttcp s1 application running on CPU3
  - Put file1/s1:
    9161.1181 Mbps 55 %TX 94 %RX 0 retrans 0.11 msRTT
  - Put file2/s2:
    8663.7400 Mbps 52 %TX 89 %RX 0 retrans 0.11 msRTT
  - Aggregate throughput:
    17824.8581 Mbps (90% of maximum 19.8 Gbps)

Introduction To GSFC High End Computing 20, 40 &100 Gbps Network Testbeds

Nuttscp Sample Test Results Between Two "C-Systems" (7-of-7) [Source: Bill Fink/GSFC]

- We are currently investigating SSD technology, to hopefully double our disk transfer speeds and get us into the 40-Gbps networked disk transfer realm

- But using parallelism of multiple cores and multiple streams is going to be key to going to 40-GigE, 100-GigE, and beyond speeds, since individual cores are not getting significantly faster