

Knots concept – yet another computing solution

- 1) knots computer is an internal network-based fat client computer architecture for GPGPU computing with practical aspects in mind. It is designed for small Research & Development up to 10 people teams. It consists of a collection of fully operational user workstations and a background distributed computing architecture,
 - 2) it is based on Linux operating system and NVidia C++11 CUDA GPU programming language provided independently of single user Personal Computer workloads,
 - 3) single knot (from overall recommended 10 knots per switch) proposition:
 MOTHERBOARD: FM2A88X Extreme6 (1xPCIe 3.0 x16; 1xPCIe 3.0 x4; ATX format),
 CPU: AMD A10 7850k (Accelerated Processing Unit – CPU with GPU),
 RAM: 2x8GB 1333MHz (majority of users works seldom exceeds 6GB of used RAM),
 HDD: 7x120GB SSD Goodram CX 300 in RAID 0 configuration,
 NET: 1port 10GbE PCIe adapter (SR SFP+ GBIC) at PCIe5 slot,
 GPGPU: user-independent **GTX 1070 8GB** at PCIe2 slot, (theoreticall ~3.4TFLOPS),
 MONITOR: 2x23.8” HP 24er connected to motherboard,
 HEADSET: Sennheiser HD280, keyboard: Lenovo Combo,
 STUFF: cool-looking case with at least 500W power supply; pendrive;
 - 4) internal 10GbE network + disk array knot
 -10GbE stacking switch (4x10GbE Cat7 RJ45 uplinks + 12x 10GbE SFP+),
 -ordinary knot without GPU + network drives on two ancient Dell MD1000 with new disks (30TB on 2TB SATA 7.2krpm; RAID 5) accessible via 8Gigabit Fibre Channel. Disk arrays are in RAID 0 configuration in total RAID 50 (hardware RAID's 5 + software RAID 0),
 - 5) internet connection with at least 300Mbps bitrate on 24ports GbE switch via 2 firewalls cascade (bought from different manufacturers). Redundant access to knot is quite useful in author opinion,
 - 6) each knot is internally interconnected via 10GbE switch. Knots are easily extendable with switches, only disk array knot should be made on server with more PCIE slots (f.e. Dell r910),
 - 7) distibuted – computations program development:
 -single GPGPU efficient kernel program development at first step,
 -data distribution from network disks to local drives via Samba. Then data partially is consecutively loaded to RAM shared memory (#shmget),
 -GPGPU works on shared memory on its host (some problems can be solved with direct reads from network to GPU memory, but author does not recommend such solution),
 -background Operating System service for shared memory handling,
 -GNU Parallel (please cite) package functionality for work deployment,
 -background service for turning off unused knots and Wake on LAN's magic packets,
 - 8) 10GbE is less problematic, much more popular and poses higher capabilities than some exotics like Infiniband and Fibre Channel network devices - but it is slower with higher connection latency,
 - 9) data distribution between knots via network is ~8 times slower than RAM shared memory access. Computational problems should be assymetric in read-write operations to computation. Please note quite acceptable connection between knots,
 - 10) there are 4ports GbE PCIe adaptors which could be connected to switch (for example 52 ports GbE) via knot 4xGbE aggregated link in more economical designs,
 - 11) please note some custom cases filled with vaseline oil providing significant heat distribution efficiency increase. Long term computations should be provided on the basis of two **GTX 1080ti**'s,
 - 12) knots concept does not require devices airconditioning resulting in lower overall system power,
 - 13) above mentioned design of 10 knots might provide practical peak of ~60TFLOPS computational capabilities for 40k\$. Annually it could consume 18k\$ of current at full load.
- Post Scriptum: additionally each knot CPU via AMD APU provides ~0.8TFLOPS computations capabilities in heterogenous programming model. Please note, that knots capabilities could be easily extended with Volta microarchitecure devices in future.