

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 CUDA GPU technology provided independently of single user Personal Computer workloads,
- 3) single knot (from overall recommended 10 knots) proposition:
 - Motherboard: FM2A88X Extreme6 (2xPCIe 3.0 x8; 1xPCIe 3.0 x4; ATX format),
 - CPU: AMD A10 7850k,
 - RAM: 2x8GB 1333MHz (majority of users processes seldom exceeds 6GB of used RAM),
 - HDD: 7x120GB SSD Goodram CX 300 in RAID 0 configuration,
 - NET: 2port 10GbE PCIe adapter (SR SFP+ GBIC),
 - GPGPU: user-independent **GTX 1070 8GB** + ordinary GT730,
 - MONITOR: 2x23.8" HP 24er connected to GT730,
 - HEADSET: Sennheiser HD280,
 - KEYBOARD: Lenovo Combo,
 - cool-looking case with at least 500W power supply; pendrive; mobile phone: myPhone Hammer, uSD card; comfortable seat,
- 4) internal 2x10GbE network + disk array knot
 - two 10GbE stacking switches (4x10GbE Cat7 RJ45 + 12x 10GbE SFP+); uplinked via copper 4 10GbE ports,
 - ordinary knot without computing GPU + network drives on two ancient MD1000 with new disks (30TB; RAID 5) accessible via server 8GFC, disk arrays in RAID 0 configuration in total RAID 50 (hardware RAID 5 + software RAID 0),
- 5) internet connection with at least 300Mbps bitrate on 24ports GbE switch via 2 firewalls cascade (bought from different manufacturers),
- 6) each knot is connected to switch via 2x 10GbE internal network. Knots are easily extendable with switches, only disk array knot should be made on server with more PCIE slots (Dell r910),
- 7) programs development:
 - single GPGPU efficient kernel program development at first step,
 - data distribution from server supervisor to RAM shared memory (#shmget) via Samba,
 - GPGPU works on shared memory on its host (some problems can be solved directly reading files 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 on server supervisor work deployment,
 - background service for turning off unused knots and Wake on LAN's magic packets,
- 8) 10GbE is less problematic and much more popular than some exotics like Infiniband and Fibre Channel network devices but yes – it is slower with higher connection latency,
- 9) data distribution via network is ~4 times slower than RAM shared memory access – computational problems should be assymetric in read-write to computation time comparisons. Please note quite acceptable connection between knots at cost of less equal two packet hops,
- 10) there are 4ports GbE PCIe adaptors which could be connected to switch (for example 52 ports GbE + 4 ports 10GbE server uplinks) 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,
- 13) above mentioned design might provide practical peak ~60TFLOPS computational capabilities for 40k\$. Annually it consumes 18k\$ of current at full load.

Post Scriptum: please note, that knots capabilities could be easily extended with Volta microarchitecture devices in future.