Performance and power implications of hardware accelerators

Saeed Iqbal

Global Solutions Engineering – HPC

Agenda

- Cluster design – "Performance chain"

- Performance characterization of an accelerator

- Sensitivity analysis of an accelerator

- Tool #1 – Compute nodes with accelerators

- Tool #2 – A first order design of cluster

- Compute nodes form factors

- Resources

Tesla K40 is the latest GPU from NVIDIA - designed for compute acceleration

- K40 has high raw compute power!
  - **4.3 – 5.4 X CPUs** (Theoretical peak)

- Compare K40 vs. K20
  - › Cores   2880   15%
  - › **Memory   12GB   240%**
  - › Mem. BW   288GB/s   38%
  - › Clock   745MHz   5.6%
  - › Power   235W   4.4%
  - › SP   4.0 TFLOPs   13%
  - › **DP   1.4 TFLOPs   20%**
  - › Kepler (GK180) architecture   (new)
  - › PCIe   Gen 3 (improved from Gen2)

Tesla K40 GPU

- Challenge:
  - How to "Realize" & "Extract" maximum real performance ?

OK, K40 GPU is a powerful.
Yes, how to get the most out of it?

GPUs →
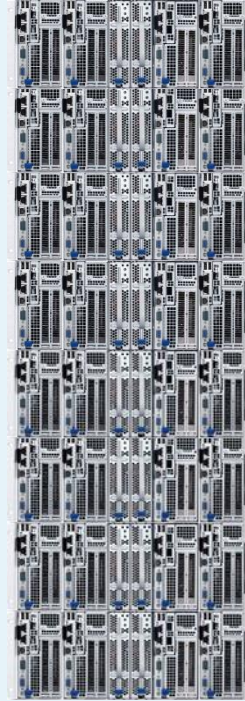
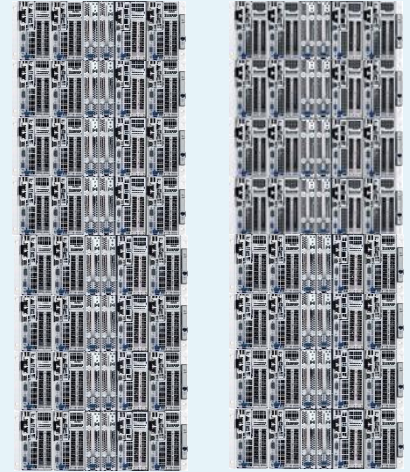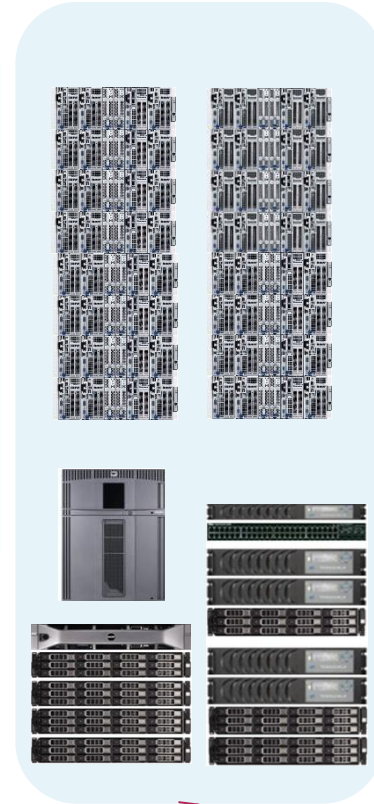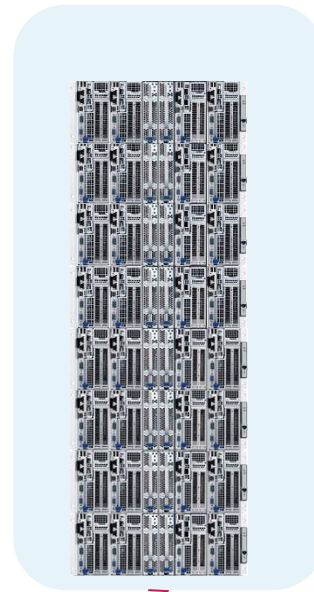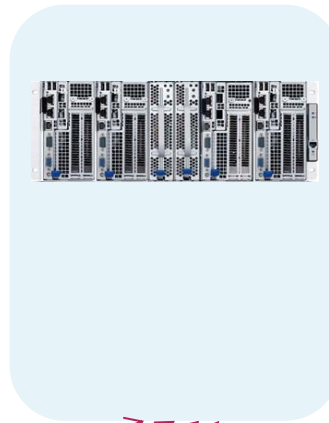**Servers →**

Clusters →

Solutions

- "Performance Chain"

- "Balance"
  - Eliminate bottle necks

- Maximize Return from Investment
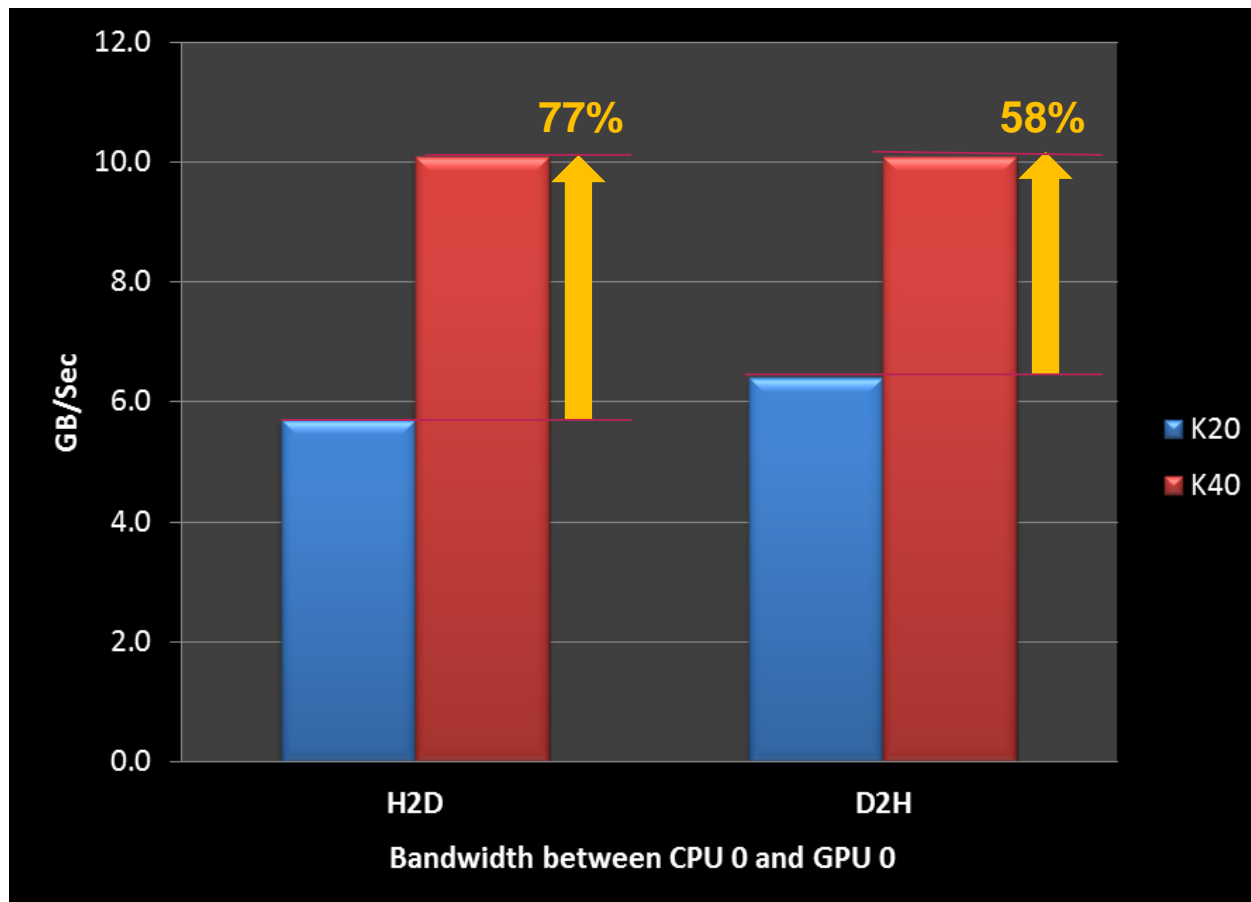
Some key issues: cluster design

- Number of GPUs per node

- Dedicated GPU nodes ?

- # GPUs nodes ?

- Local memory/storage

- Power budget per rack

- Number of nodes with GPUs for best ROI

Performance characterization of an accelerator
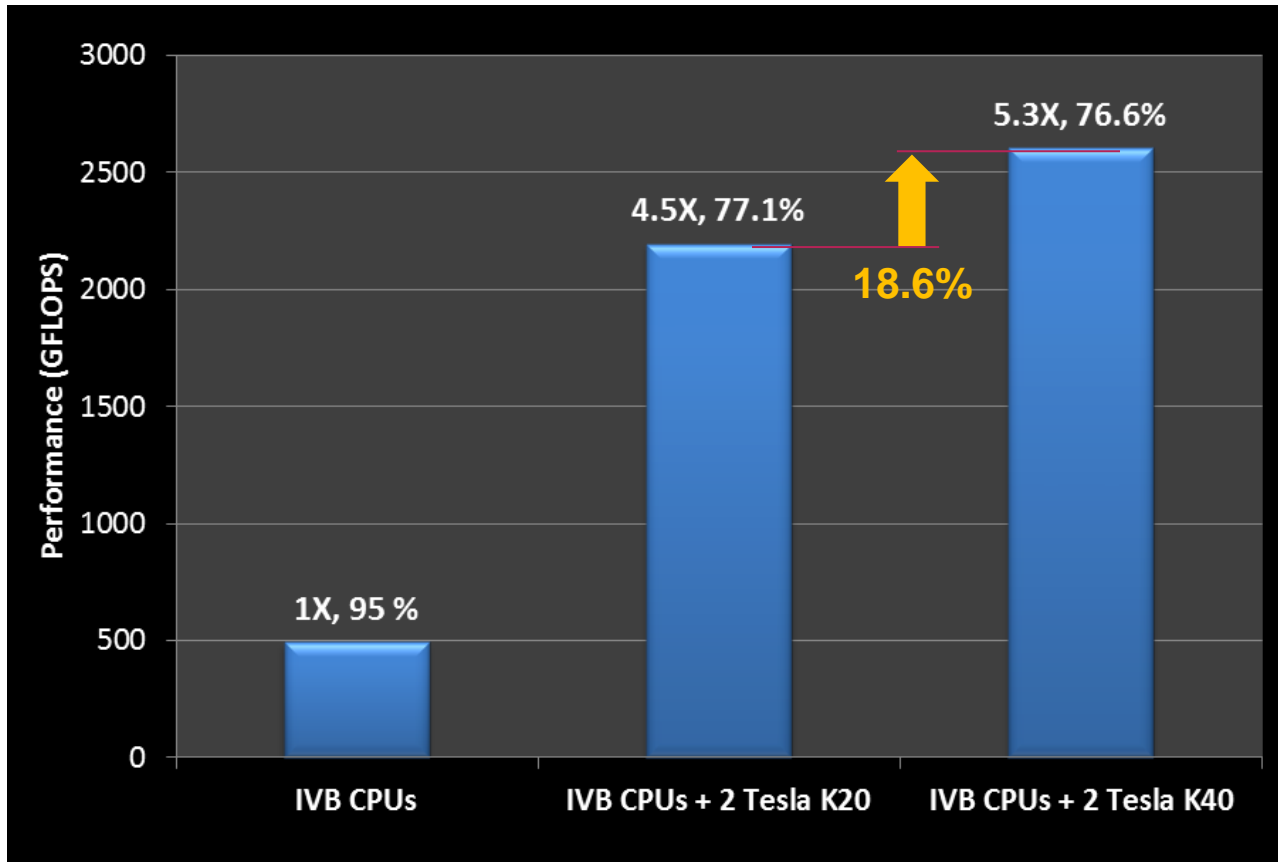
K20 vs. K40 – Bandwidths (Pinned) R720 with E5-2697 v2 CPUs, CUDA SDK BW

K40 improves the H2D BW by 77% & D2H BW by 58%
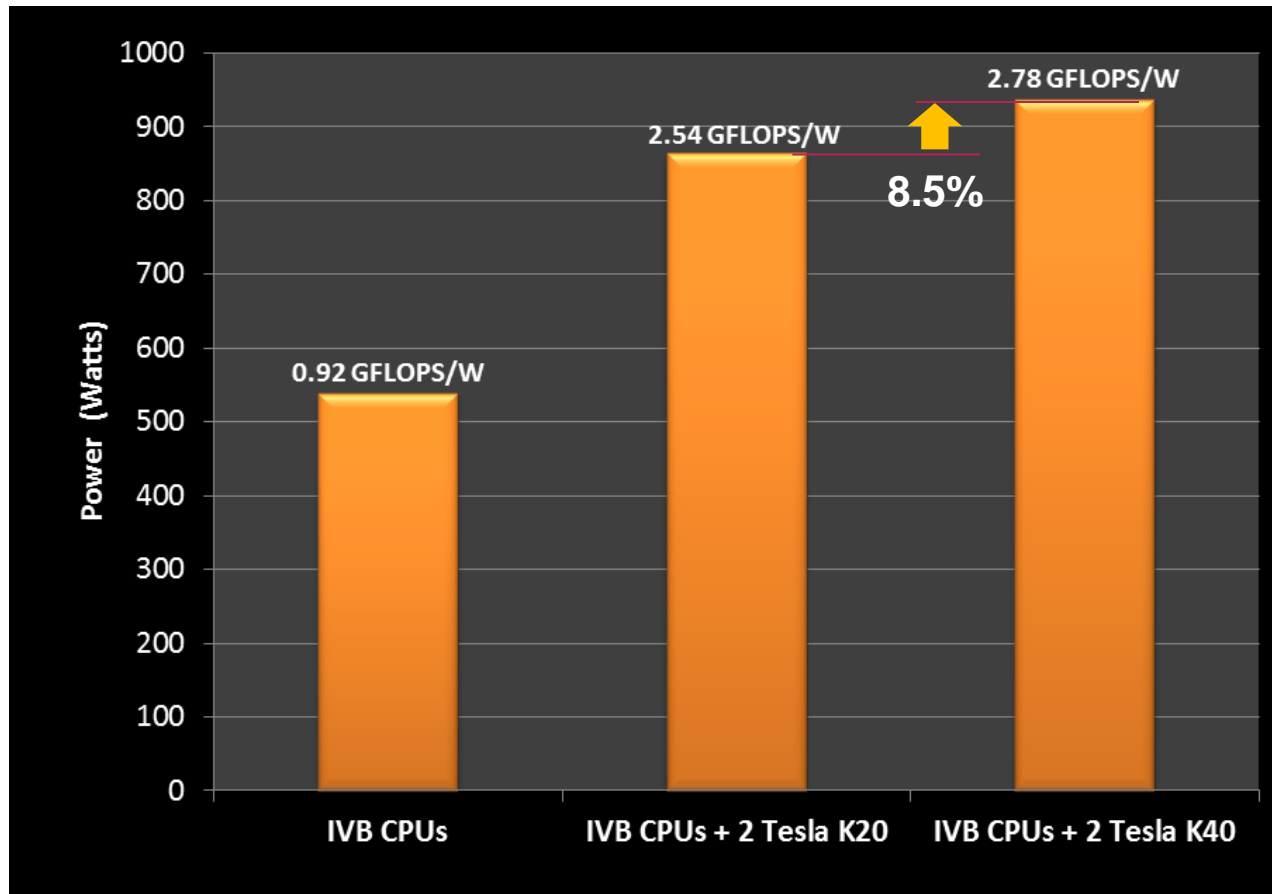The improved BWs will improve application performance!

PE R720, Dual E5-2670 v2@2.6GHz, 128GB 1600MHz memory, Tesla K20 & K40, CUDA 5.5 (319.49), CUDA BW (numactl -c0/1, -m0/1)

K20 vs. K40 – HPL Perf. & Eff. R720 with E5-2697 v2 CPUs

K40 improves performance by 18.6% and acceleration 5.3X !

HPC Engineering

K20 vs. K40 – HPL Power R720 with CPUs

K40 improves GFLOPS/w by 9.2%.
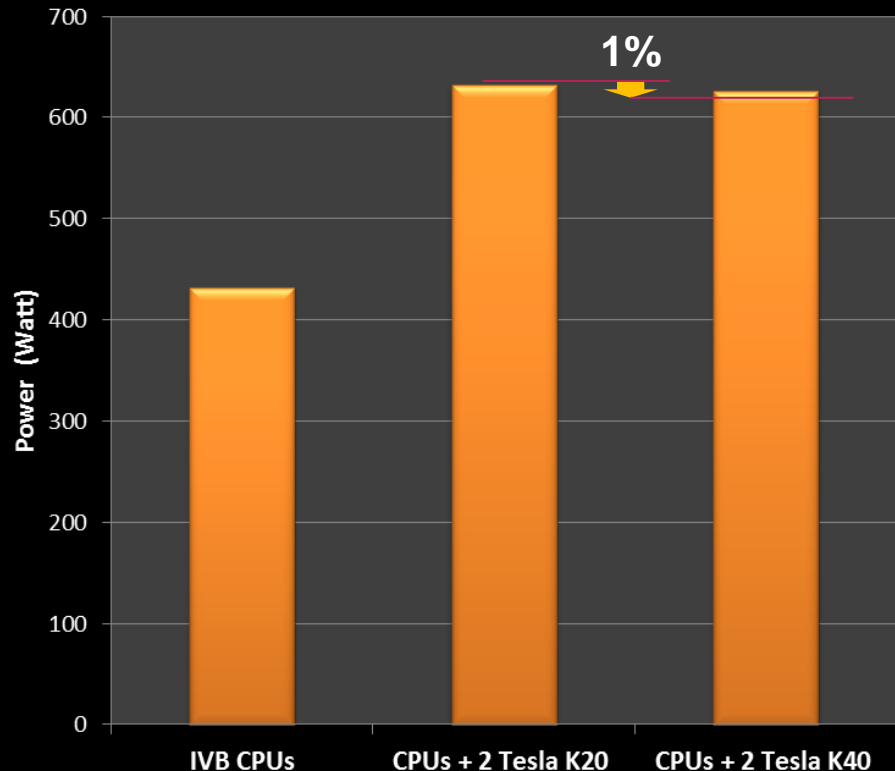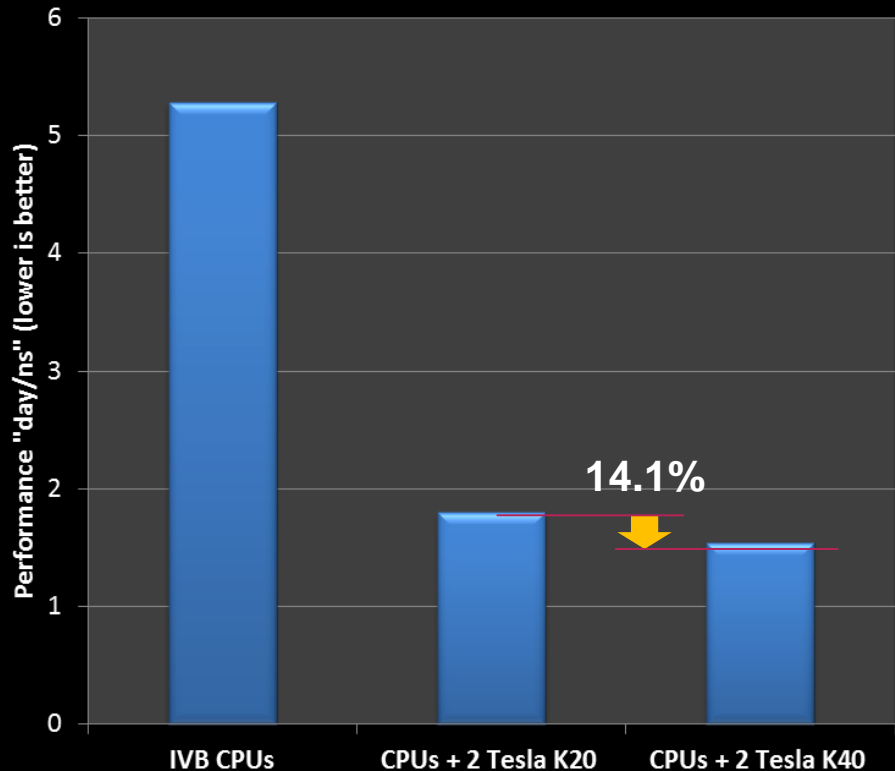**Power consumption increases by 8.5%. to 3.0X CPU-only system.**

PE R720, Dual E5-2670 v2@2.6GHz, 128GB 1600MHz memory, Tesla K20 & K40, CUDA 5.5 (319.49), nVIDIA HPL 2.1

K20 vs. K40 – NAMD
R720 with E5-2697v2 CPUs, STMV (1M atoms)

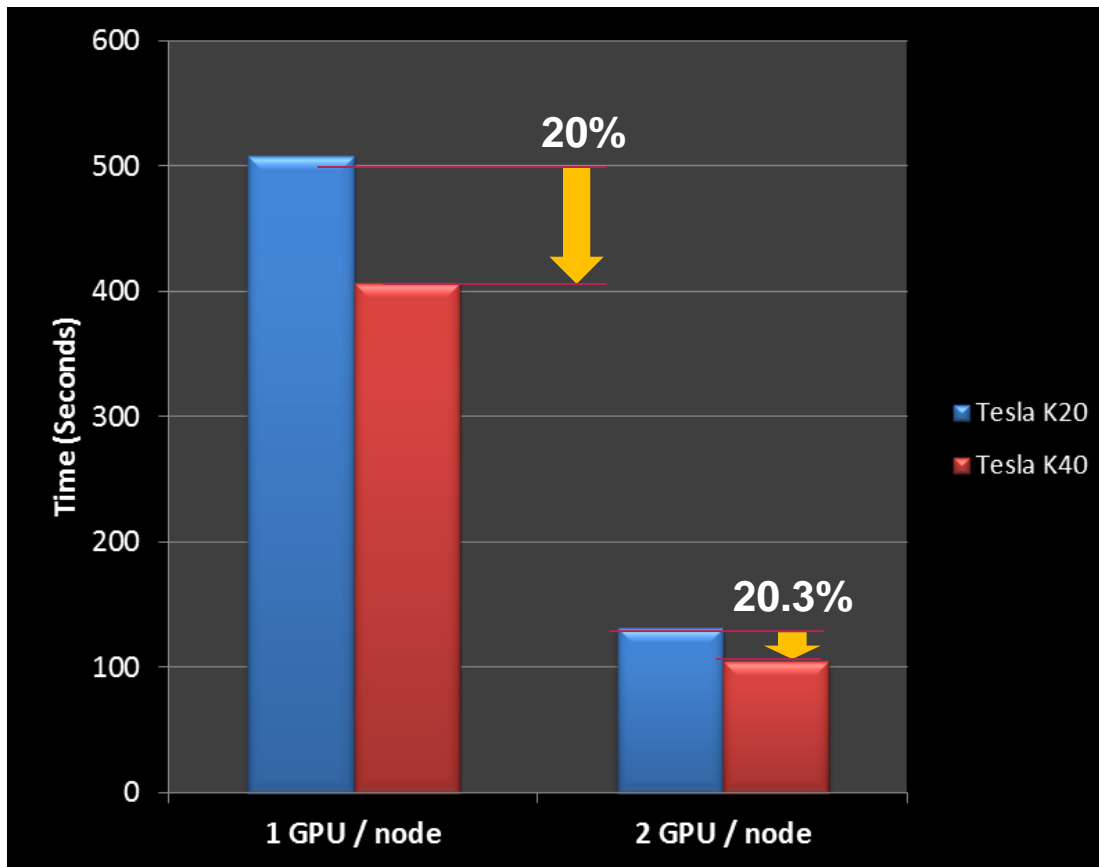**K40 improves NAMD performance by 14.1% for large simulations; at about the same total power consumption!**

PE R720, Dual E5-2670 v2@2.6GHz, 128GB 1600MHz memory, Tesla K20 & K40, CUDA 4.2, NAMD 2.9

HPC Engineering

K20 vs. K40 – NBODY
R720 with E5-2697v2 CPUs, (N=1000000)

K40 improves NBODY performance by 20% for large simulations.
There is a 3.9X acceleration due to the second GPU !

PE R720, Dual E5-2670 v2@2.6GHz, 128GB 1600MHz memory, Tesla K20 & K40, CUDA 5.5 (319.49), NBODY CUDA SDK
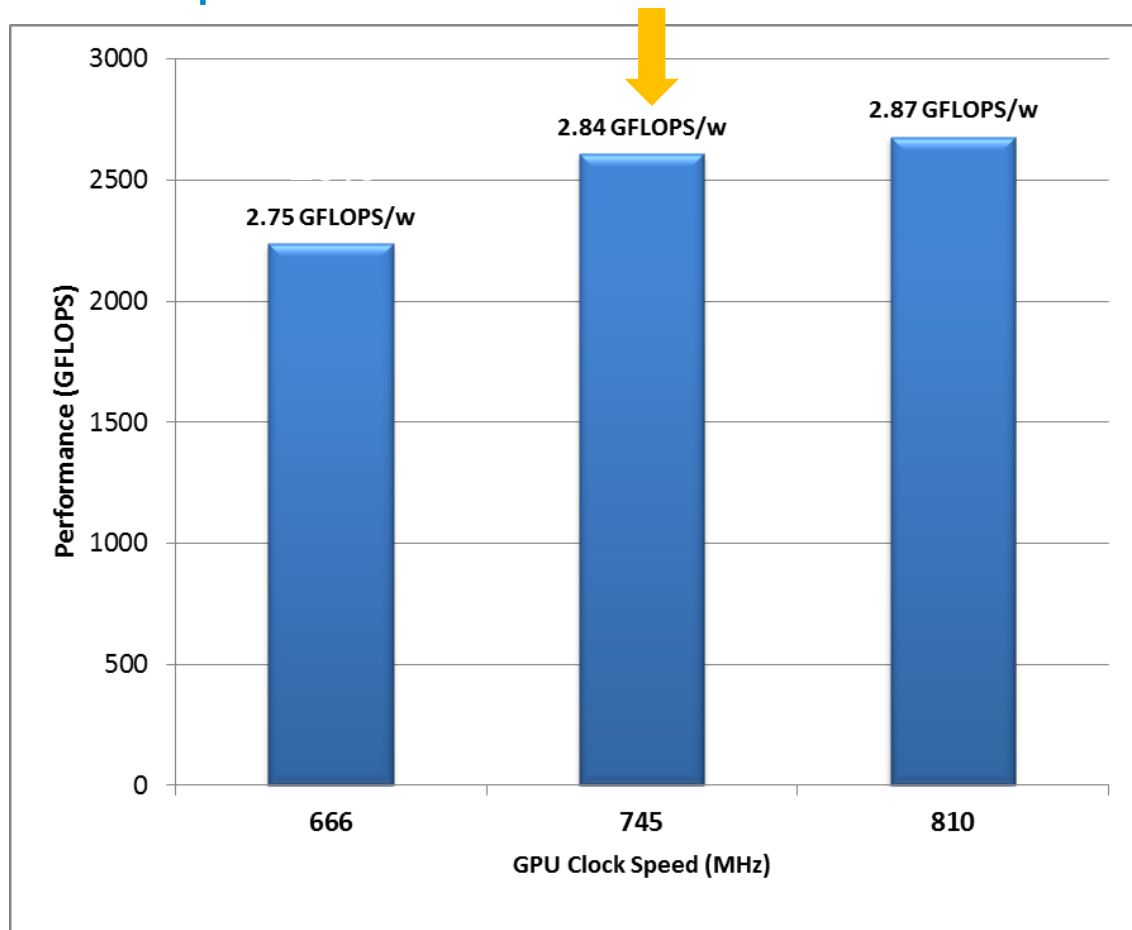
Summary of characterization results
(K40 vs. K20)

- K40 shows 18.6% better HPL performance

- K40 requires **8.5%** more power for HPL

- K40 shows **9.2%** improvement on HPL GFLOPS/watt

- K40 has up to **14%** improvement on STMV acceleration
  (1 million atom benchmark of the NAMD)

- K40 has up to **20%** improvement for NBODY simulations
  (N=1000000)

K40 can improve performance by 15-20% for about 10% more power, compared to K20.
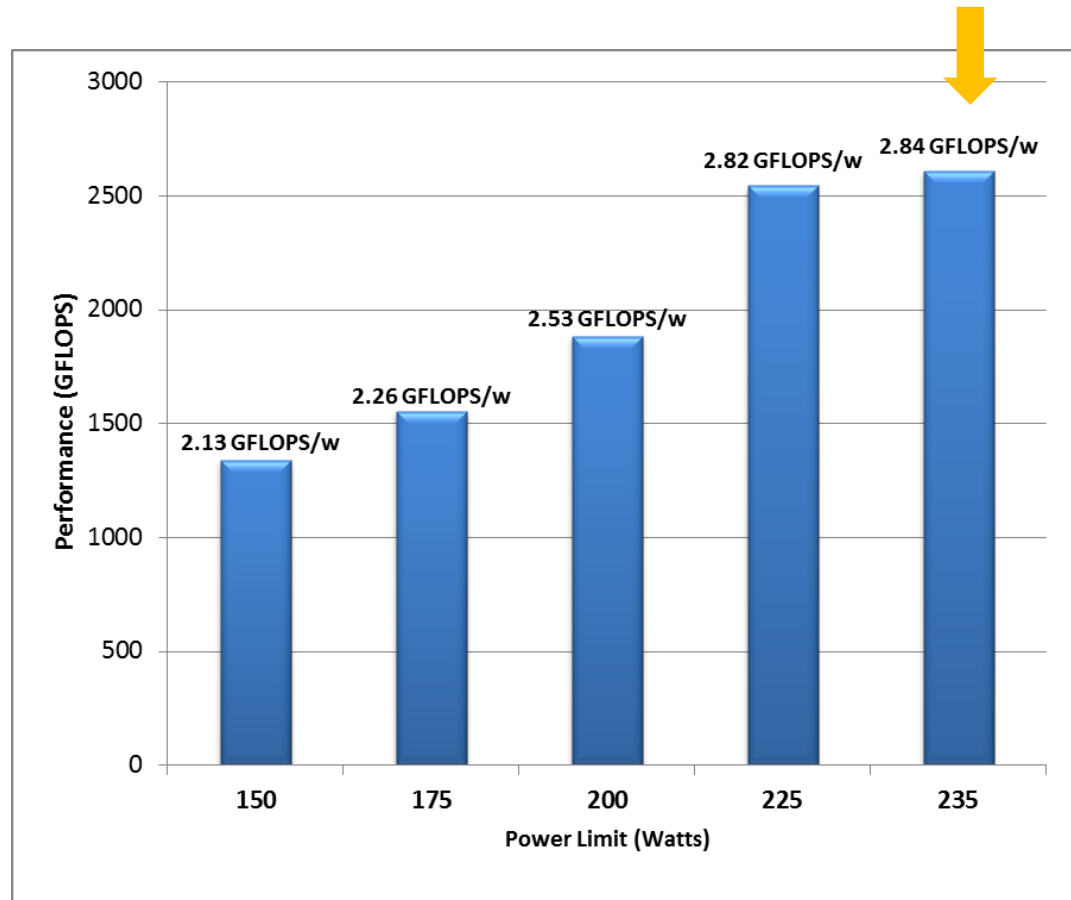
Sensitivity analysis of GPU performance

K40 parameter sensitivity – Changing "power limit" and "GPU clock speed"

- On a K40, power consumption and clock rate can be adjusted:
  - GPU Clock Speed options → [666, **745**, 810, 845 MHz]
    - › nvidia-smi -ac --application-clocks=<memory, graphics>
  - GPU Power limit options → [**235**, 225, 200, 175, 150 W]
    - › nvidia-smi -pl --power-limit=<limit>

K40 Performance sensitivity GPU clock speed at 235W

K40 performance improves 2.7% due to overclocking to 810MHz

PE R720, Dual E5-2670 v2@2.6GHz, 128GB 1600MHz memory, Tesla K20 & K40, CUDA 5.5 (319.49), NAMD

K40 performance sensitivity
GPU power limit at 745 MHz

System power varies from 626W to 916W due to power limit
Operating at power limit of 225 can result in saving of 15W/node

PE R720, Dual E5-2670 v2@2.6GHz, 128GB 1600MHz memory, Tesla K20 & K40, CUDA 5.5 (319.49), NAMD

Summary of sensitivity analysis
K40

- K40 is more sensitive to "power limit" compared to "GPU clock"

- Each application can have its own optimal setting of GPU parameters

- These setting offer different "operating points" for accelerators

HPC Engineering

Tool #1

What percentage of cluster nodes should have GPUs ?

How many GPUs nodes are cost efficient ? NVIDIA/Del has a tool to answer the question

- Scenarios
  - Current customers may be running GPU enabled apps on CPU only systems.
  - New customers need help maximizing ROI.

- Inputs
  - **Fixed Budget**
  - Should I buy ? (two options)
    - › **CPU-only node**
    - › **CPU+GPU node**
  - **Goal: Maximize job throughput**
    - › Future change in application mix
    - › Power Savings due to GPUs

How many GPUs nodes are cost efficient ? NVIDIA has a tool to answer the question
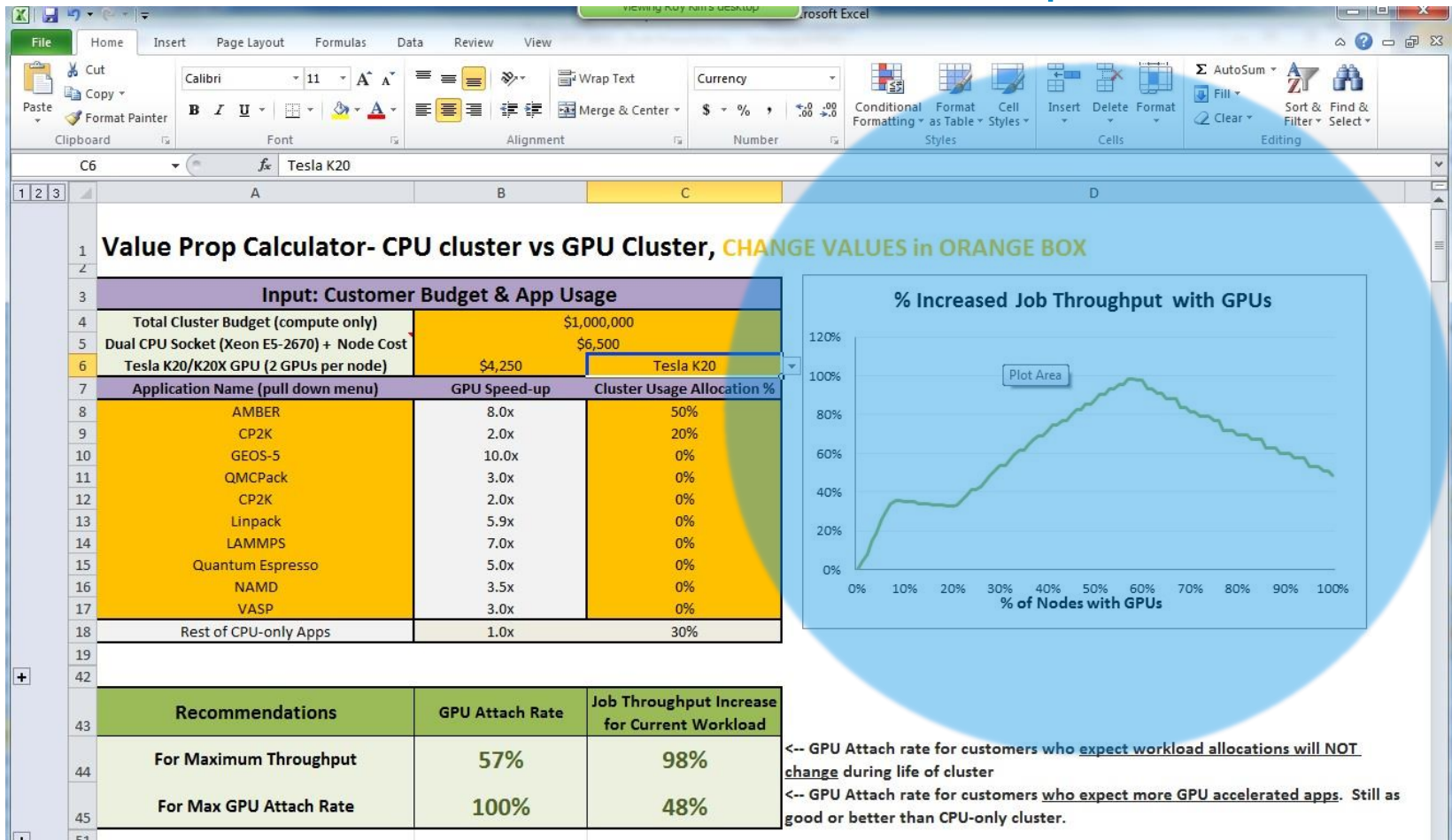
Outputs: **Given a Budget & "Application Mix", how to maximize job throughput**

Note: Please contact Dell Sales Contacts (Tool is not publically available).

Tool #2

I want to outline the design of my cluster ?

HPC Advisor Tool - Design your GPU cluster in minutes!

- Public software application that recommends a Dell HPC solution based on customers specific needs

- **Goal: Create Balanced Cluster Designs**

- Example: The HPC Advisor asks user:

  - OS type? **GPU**? Server Form Factor?

  - Optimize for performance, power or density

  - Desired sustained or theoretical performance

  - Recommends a solution based on this input.

Available on Dell.com http://dell.com/hpc
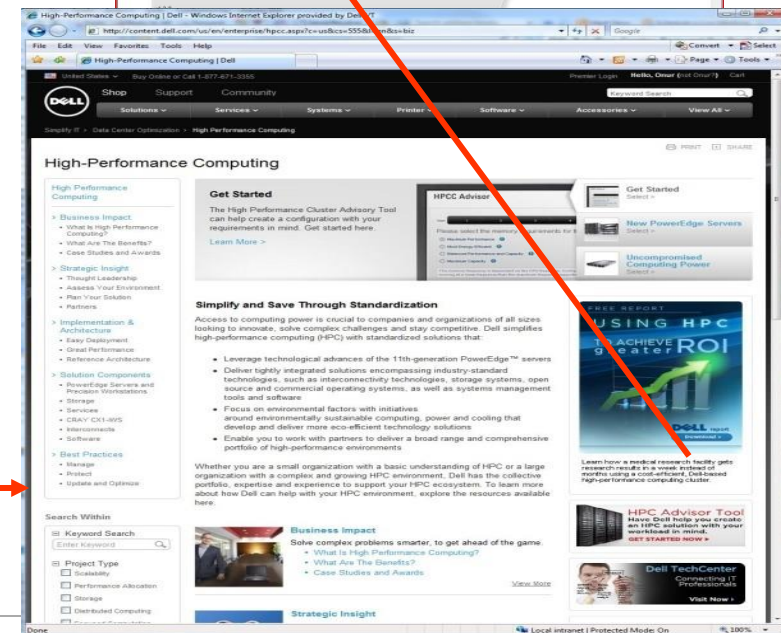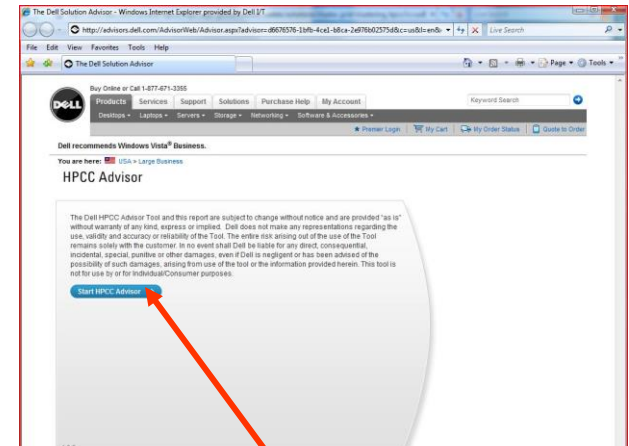
Resources

where to get more information, blogs, etc.

Resources

- Blogs

- Whitepapers

- www.dell.com/gpu
- www.dell.com/hpc
- www.hpcatdell.com
- www.DellHPCSolutions.com
- http://www.hpcadvisorycouncil.com/best_practices.php

Resources: www.dell.com/gpu

- Overview

- Supported GPUs

- GPU Specs

- GPU Solutions

GPU accelerators and coprocessors for PowerEdge servers

Hundreds of cores for incredible performance.
Add graphics processing units to your PowerEdge servers for increased processing power.

| Overview | GPU Options | GPU Specs and Support | Solutions |

Resources

GPU-Accelerated Solutions

HPC Solutions Advisor

HPC Tech Center

HPC GPU Tech Center

GPU accelerators are available for the PowerEdge R720, T320, T420, T620 and C8220x servers and the C410x PCIe expansion chassis.

Increase the performance of your PowerEdge data center.

Extract some of the highest levels of performance from your Dell PowerEdge servers through a general-purpose computation on graphics processing units (GPU) architecture. When you add GPU processing power to the CPU capabilities already available in your PowerEdge servers, you open the door to outstanding performance across hundreds of processing cores.

- GPUs are high-performance, many-core processors that can be used to accelerate a wide range of applications.
- Advanced GPU programming methods and toolkits enable easy integration into your data center.
- GPU processors can be internally installed in standard PCIe slots or connected externally via

Compute nodes form factors

Two Server Form Factor Options Ready for K40/K20 GPUs

- ## PowerEdge C8220X
  - "Shared Infrastructure"
  - 4U
  - Higher GPU & CPU Density
  - Higher Configurability

- ## PowerEdge R720/R730
  - "Conventional Rack Server"
  - 2U
  - Higher memory per node (768GB)
  - Higher storage per node (24TB)

C8000+C8220X

R720

3D View

3D View

The C8000 Series: CPU, CPU+GPU Sleds

Based on the "Shared Infrastructure" design

C8000

C8220   C8220X   C8220XD

- C8220 (single wide, compute sled)
- C8220X (double wide, compute sled)
- C8220XD (double wide, storage sled)

As demands change it can be reconfigured or scaled out extending the life and value of IT infrastructure investments
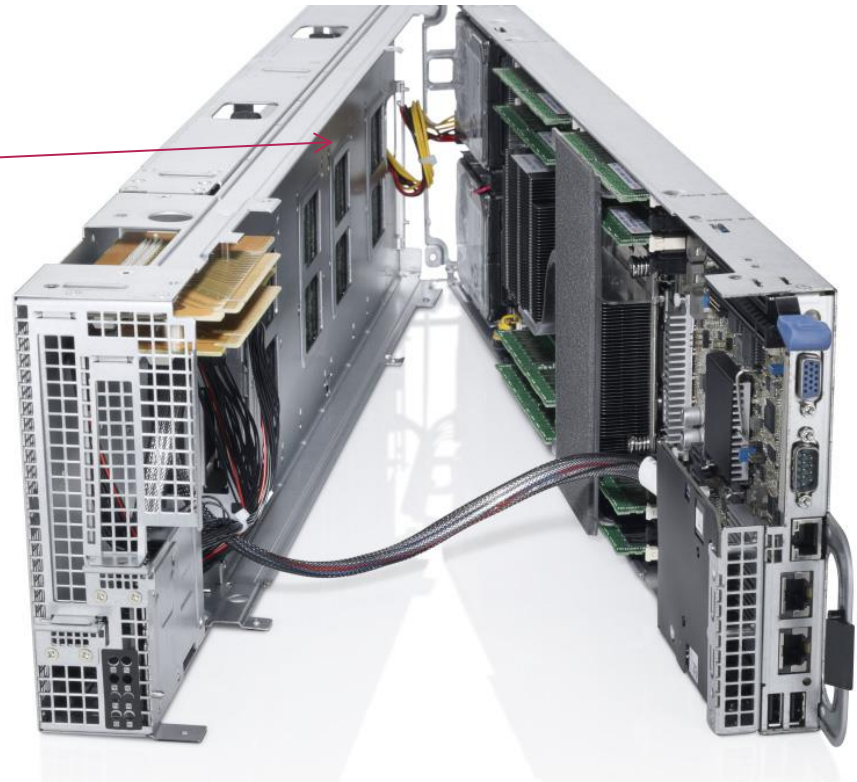
HPC Engineering

Server Details: PowerEdge C8220X

Each C8220X has:

- Up to 2 K20 GPUs

- Two E5-2600 CPUs

- 256GB of memory

- Combine sleds
  - **4 C8220X Sleds in one C8000**

- 8 GPUs in 4U space
  - **2 GPU/U Density**

PowerEdge
C8220X Compute
GPU Sled

# Server Details: PowerEdge R720

Each PE R720 has:

- Up to 2 K20 GPUs

- Two E5-2600 CPUs

- 768GB of memory
  - 24 X 32G DIMM

- 24TB local storage
  - 16 X 2.5TB Drives

- 2 GPUs in 2U
  - **1 GPU/U density**

PowerEdge R720

"Dell HPC Solutions" Mean "Value" for you

- Solutions "Goal" is to provide "value"
  - Enables you to focus on you "science"
  - Brings your HW up to speed *FAST*
- Engineering Rigor
  - Performance Envelop
  - Measure Total Power Consumption,
  - Expected Power efficiency
- Best practices
  - HPC Advisor
  - Whitepaper Publications,
  - Public Results

256-Node HPCC Bundle
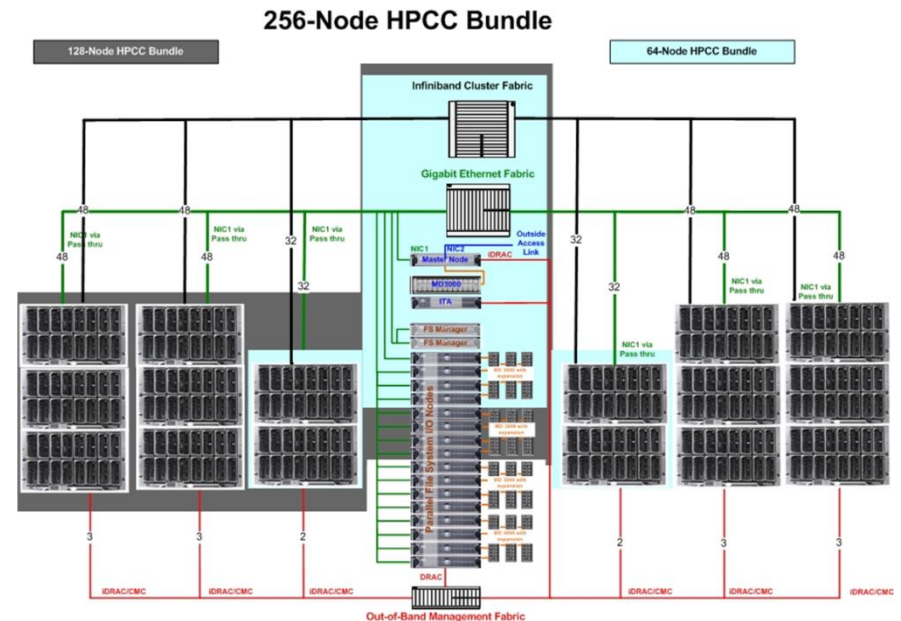
"Dell Solutions" Mean "Value" for you
Big part of it the Engineering

- Tests suite includes
  - Node level Performance
  - Cluster level Performance

- Power
  - Total Measured System Power Consumption
  - Performance/watt studies for efficient configurations

- System level
  - Host-to-device, Device-to-host, Device-to-Device
  - Memory subsystem

- Applications level : Benchmarks and Applications
  - HPL, NAMD, NPB, ANSYS

Summary of the Key Features of HPC Solutions

- **Balanced**   (GPUs, Compute, Storage, Networking)

- **Powerful**   (1 or 2 GPU/U Density)

- **Adaptable**   (workload based configuration)

- **Flexible**   (modular components)

- **Scalable**   (Modular building blocks)

- **Efficient**   (Compared to equivalent CPU only clusters)

Start Small,

Grow and Adapt your HPC solution based on your needs!

Q&A

Thank you.

Dell HPC lab in Austin.
You are Welcome to visit us!

10/3/2014