Power, Cooling and Hardware Considerations for Next Generation Data Center & HPC Solutions

Kanti Bhabuthmal
Marketing Director
Super Micro Computer, Inc.

HPC Advisory Council
Stanford Conference 2015
February 2-3, 2015
Munger Conference Center/ Paul Brest Hall
Stanford University
# AGENDA

## HPC Cluster – Power, Cooling and Hardware Considerations

### HPC Hardware Requirements

<table>
<thead>
<tr>
<th>Primary Drivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trends</td>
</tr>
<tr>
<td>Emerging Technologies</td>
</tr>
</tbody>
</table>

### HPC Cooling Solutions

<table>
<thead>
<tr>
<th>Liquid Cooling Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supermicro Success Stories</td>
</tr>
</tbody>
</table>

### HPC Server Product Portfolio

### Summary
HPC HW Requirements - Primary Drivers

- **POWER**
- **PERFORMANCE**
- **PRICE**
- **INTERCONNECT**
- **RACK & FLOOR SPACE**
HPC – HW Drivers / Trends
HPC – HW Drivers Emerging Technologies

- Multi Node Servers
- Titanium Power Supplies
- UlltraDIMMM Memory
- NVMe Storage
- Battery Backup
  No UPS
- Liquid Cooling
Cooling Solutions
HPC Success Stories

Traditional Chilled Water Door System

Direct To Chip Heat Exchanger

Liquid Immersion
Airflow & Containment

Hot or Cold Aisle Containment
Direct the airflow of either cold aisle or hot aisle to reduce "bypass air" and "recirculation" issues.
Chilled Water Door Cooling System

Modular Designs
Systems operate in much higher Ambient

Chilled water Compressor
PNNL: Pacific North National LAB

Chilled Water Door Cooling System

Ranked #18 on Top500 Nov 2014

Requirement
• 3.4 Petaflop theoretical peak processing speed
• 2.7 Petabyte shared parallel file system
• (60GB/s read/write)

Configuration
• 1,440 compute nodes with 2,880 Intel Xeon Phi 5110P accelerators
• 128GB memory per node
• FDR InfiniBand network
• 42 racks/195,840 cores
• Liquid cooling rack rear door


78.6°F
Liquid Cooling - DCLC™ Rack Subsystem AHx

Cooling subsystem - components used in the Cherry Creek project

**AHx Module**
- Dual redundant fans
- Centralize pumping architecture
- CoolIT Command Center
  Monitors/alerts on health of liquid system

**Manifold Module**
- Steel body attaches similar to PDU
- All-metal dry-break quick connects

**Server Module**
- Passive cold plate technology
- All-metal dry-break quick connects

FatTwin™ node
**Liquid Cooling - DCLC™ Rack High Capacity CHx**

### Centralized Cooling Pump Support Multiple High Density Racks

**Features**
- Liquid-to-liquid heat exchanger
- Redundant centralized pumps
- Retrofittable
- CoolIT Control System
- Supports 10 High Density Racks

**Benefits**
- Supports 50kW of cooling per rack
- Eliminate or reduce need for chillers and CRAC’s
- Power savings allows for fast ROI
- Decreases PUE in legacy data centers
- Monitor system health remotely

**High Capacity Rack CHx**

<table>
<thead>
<tr>
<th>Feature</th>
<th>CHx – HC</th>
</tr>
</thead>
<tbody>
<tr>
<td>kW Capacity (@ 30°C facility water)</td>
<td>750kW</td>
</tr>
<tr>
<td>Racks per system</td>
<td>10</td>
</tr>
<tr>
<td>Power Consumption (Max.)</td>
<td>4.2 kW</td>
</tr>
<tr>
<td>Centralized Pump Capacity (Max.)</td>
<td>360 lpm</td>
</tr>
</tbody>
</table>

**Dry Coolers – No chilling required**

Can be installed under raised floor or above racks

**Multiple Rack CHx**
Supercomputing Cluster - Cherry Creek

Each 42u rack Installed with:
- Supermicro 48 FatTwin™ Servers - 4U/4 DP Nodes
  - 2 Xeon CPU and 3 Xeon Phi Per Node
- Liquid Cooling (CoolIT)
  - Cooling plates go in the servers for both CPUs and Xeon Phi
  - CPUs & Xeon Phi Heat Sinks removed

Results:
- 76% reduction in fan power
  - 3,600 RPM fans at 100% CPU utilization running LINPACK
  - 15,000 RPM standard with air
  - 150W savings
- 9 C rise across server (inlet to outlet)
- CPU temperature 59 C at 100% CPU utilization
- DIMM’s running at 52 C

CoolIT’s Rack DCLC™ technology
- 1 Rack DCLC AHx Module
- 2 Rack DCLC Manifold Module
- Not taking any U space
Immersion Liquid Cooling

Metal plate – 2 boards per plate
2x Truescale IB per board

Dielectric Mineral Oil Blend as a Coolant
Non Electric but Good Heat Conductor

Intel Truescale IB Switches
Supermicro Ethernet Switches
Supermicro 2 DP Server Boards

Rack liquid tank top view

PDU Mounts on front or back of rack

Vertically mounted OEM servers
Power cable guides
Liquid fillline

Coolant returning to the rack from heat exchanger at user-specified temperature

Heated coolant exits at top of rack

Coolant-to-water heat exchanger

Coolant pump

86° - 127° F Water Source

Numerous options for cable organization and routing

Vertically mounted OEM servers

Dielectric coolant inlet (exit not pictured)
### Vienna Scientific Cluster VSC-3

Largest HPC Deployment from Supermicro 2020 nodes with Submersion Liquid cooling

| **Motherboard** | Supermicro X9DRD-IF  
|                 | Intel Patsburg Chipset  
|                 | QuickPathInterconnect (QPI) 8.0 GT/s  
|                 | Dual Xeon Sandybridge (E5 Series)  
|                 | Up to 256GB DDR3 1600/1333/1066/800MHz ECC Registered DIMM in 8 DIMM sockets  
|                 | Expansion slots: 1 (x16) PCI-E 3.0 and 4 (x8) PCI-E 3.0 slots  
|                 | Intel® 350 Dual-Port Gigabit Ethernet Controller  
|                 | 6x SATA2 and 2x SATA3 ports  
|                 | Integrated IPMI 2.0 with Dedicated LAN  
|                 | Supermicro RSC-RR1U-E8 1U PCI-E Riser Card |

| **Chassis** | SNK-P0047P passive 1U heat sink for X9 Generation Motherboard  
|             | Indium Foil replaces heatsink paste  
|             | 1U PowerSupply ~350W |

| **CPU** | 2 x Intel Xeon IvyBridge-EP E5-2650v2  
|         | 2.50GHz  
|         | 8 Core - 20MB Cache  
|         | Intel HT Technology - Intel Turbo Boost Technology  
|         | 95W TDP (Thermal Design Power) |

| **Memory** | 8 x 8192MB DDRIII1866 ECC Registered (512Mx8) |
Submerged Supermicro Servers Accelerated by GPUs

- 1U dual-GPU SuperServer for Geoscience / Oil & Gas applications
- No requirement for room-level cooling
- Operates at PUE ~ 1.12
- 25 kilowatts per rack – the breakpoint = 20kW ~ 25kW per rack (between regular air-cool and submerged cool)
- Full warranty: [http://www.grcooling.com/about/blog/](http://www.grcooling.com/about/blog/)

“The 24-rack installation at CGG Veritas in Houston is the largest deployment to date of Green Revolution’s CarnotJet cooling racks, which are filled with 250 gallons of dielectric fluid,.....”
Submerged Cooling Solution Breakthrough

Green500 #1 on Nov., 2013 List

- PUE ~ 1.05
- > 3GFLOPS / watt

NEC/SMIC 1U Server x 40 Nodes Each node:
- 2x Intel® Ivy-Bridge 2.13GHz 6-Core
- 4x NVIDIA Tesla K20X GPU
- 543GB DDR3 memory
- 120GB SSD
- 4x FDR InfiniBand 56Gbps
- Total Peak: 210TFlops (GP); 630TFlops (SP)


- Fluid Submersion Cooling +
  Outdoor Air Cooling +
- High Density GPU Supercomputing in a 20-feet container (15m²)
- Cooling Tower: Water 25-35°C >> Outdoor air

- World's top power efficiency (>4.5GFlops/Watt)
- Average PUE 1.05, lower component power
- Field test ULP-HPC results, TSUBAME3.0 Prototype
Server Technology Innovation
Supermicro Ultra Servers

- Dual Xeon Processors E5-2600v3 up to 160W
- 8 PCIe Expansion Slots (1 PCIe x16 FHFL, 5 PCIe x8 FHFL, 1 PCIe x8 LP, 1 PCIe x8 LP Internal)
- Redundant 1000W Titanium PSU
- 24 DIMM Support
- 24 2.5” Hot Swap Bays

- Dual Xeon Processors E5-2600v3 up to 160W
- 4 PCIe Expansion Slots (2 PCIe x16 FHFL, 1 PCIe x8 LP, 1 PCIe x8 LP Internal)
- Redundant 750W Platinum PSU
- 24 DIMM Support
- 10 2.5” Hot Swap Bays
Multi Node Servers - MicroBlade

Intel™ Atom C2000
112 Nodes

MBE-628L-816/416

Intel™ Xeon UP E3-1200 v3
28 Nodes

MBE-628E-816/416
MBE-628E-420

X10 Xeon DP E5-2600 v3
28 Nodes

MBE-628E-816/416
MBE-628E-820/420

Built in Switching Modules. Eliminates all cabling. Optimized for Small workloads
Multi Node Servers - MicroCloud

SYS-5038ML-H24TRF
- 3U 24 nodes
- Intel C224 PCH Chipset
- Intel Haswell CPU support from 13 Watt up to 80Watt
- 32GB VLP DDR3 ECC UDIMM
- 2 x 2.5" SATA HDD or 4 x 2.5" SSD per node
- 2000 watt redundant power

SYS-5038ML-H12TRF
- 3U 12 nodes
- Intel Haswell CPU support up to 80Watt
- Intel C224 PCH Chipset
- 32GB VLP DDR3 UDIMM
- 2 x 3.5" or 4 x 2.5" SATA HDD per node
- MicroLP upgradable
- 1620Watt Redundant / (2000Watt optional)

SYS-5038MR-H8TRF
- 3U 8 nodes
- Intel Haswell EP
- 512GB DDR4 ECC RDIMM/LRDIMM
- 2 x 3.5" Hot-Swap SATA HDD
- MicroLP upgradable
- 1x PCIEx8 low profile card support per node
- 1620 watt redundant power

24 UP Nodes
E3-1200 V3

12 UP Nodes
E3-1200 V3

8 DP Nodes
E5-2600 V3

1U Rack Server Vs. Microcloud = 72% Space Saving  70% Fewer Cables  27% Power Saving
Summary
Performance and Energy Efficiency

Energy Efficiency

HISTORICAL CONTEXT

2014, < 0.33 MWatts/PFLOP

ExaFLOPS ~2020

Performance

EFLOPs

PFLOps

TFLOPs

GFLOPs

SCFLOps

SMP / MPP
Proprietary Solutions

Commodity Components
General Purpose, off the shelf,
PC cluster

App. Optimized
Blades or High Density, High
Efficiency Servers

Scalability and
Performance / $$$

TeraFLOPs
Challenge

GigaFLOPs
Challenge

10K Watts /TFLOP
Challenge

PetaFLOPs
Challenge

Efficiency & Density
Performance / Watt / FT²

Hybrid System
CPU + GPU

Efficiency & Density
Performance / Watt / FT²
Thank You