Capitalize on Next Generation In-Memory HPC with HPE Superdome Flex

Bill Dunmire
HPC Solutions Marketing, HPE

March, 2018
What do these compute problems have in common?

- Challenging to distribute across multiple nodes in an HPC cluster. Want to **solve holistically**.
- Compete for “fat nodes” (with more processors/memory), **compounding time and effort to discovery**
- Or **too big** for the node
  - If memory is exhausted, simulation fails (hours wasted)
  - Use smaller models, reducing accuracy and increasing prototype costs/time to production
Do you have a dedicated cluster administrator?

Small distributed HPC environments

...or tasking engineers and scientists to learn & manage, taking time away from development and research?
Solution: HPE In-memory High Performance Computing
Advanced SMP systems provide globally shared memory with seamless scale up capacity

- Solve complex, data-intensive problems **holistically** at **unparalleled** scale with **single-system** simplicity
- Complete **more jobs in less time**
- Free HPC teams from managing clusters to **accelerate time to discovery**

**Use cases include:**
- Genomics
- Bioinformatics
- Computer-aided engineering
- Computational chemistry
- Cyber security
- Financial risk management
- Fraud detection and prevention
- Real-time, streaming graphs
- Large data visualization
- In-memory databases
HPE Superdome Flex: Innovation platform for the next decade
Combining the best of HPE and SGI technology for next generation in-memory HPC

**HPE Superdome X**

- **Scale**: Bladed scale-up architecture
  2-16s (Broadwell), up to 48TB shared memory

- **IO**: Mezz cards, 24 max +16 FLBs
  External drives only

- **Flexibility**: HPE nPARs, VMware, KVM
  Higher cost entry point

- **Availability**: Extreme RAS, self-healing

- **User Experience**: HPE OneView management

**HPE MC990 X (renaming of SGI UV 300)**

- **Scale**: Modular scale-up architecture
  4-32s (Broadwell), up to 48TB shared memory

- **IO**: PCIe standup cards, 96 max

- **Flexibility**: Lower cost entry point, KVM

- **Availability**: High RAS

- **User Experience**: Linux CLI

---

**HPE Superdome Flex**

- Unparalleled Scale
- Unbounded IO
- Optimum Flexibility
- Extreme Availability
- Simplified User Experience
Flexible modularity and extreme scale for your Linux HPC workloads
HPE Superdome Flex

Unparalleled Scale
- Modular scale-up architecture
- High bandwidth (7.4GB/sec)/low latency (<400ns) HPE fabric
- Scales seamlessly from 4s to 32s as a single system
- Designed to provide 768GB-48TB of shared memory
- Intel Skylake (2x FLOPs, higher mem bandwidth/lower latency, up to 28 cores)

Unbounded I/O
- Up to 128 PCIe standup cards, LP/FH PCIe

Optimum Flexibility
- 4-socket chassis building blocks, low entry cost; HPE nPARs*
- Nvidia GPUs, Intel SDVis
- 1/10/25 Gbe, 16GbFC, IB EDR/Ethernet 100gb, Omni-Path
- SAS, Multi-Rail L.Net for Lustre; NVMe SSD
- MPI, OpenMP

Extreme Availability
- Advanced memory resilience, Firmware First, diag engine, self-healing
- HPE Serviceguard for Linux

Simplified User Experience
- HPE OneView management*, IRS, Openstack
- HPE Proactive Care

* Mid 2018
HPE purpose-built portfolio for HPC

HPC and AI Industry Solutions

Financial Services
Academia, Research, Gov't
Life Sciences, Health
EDA / CAE Manufacturing
Oil and Gas, Energy
Weather and Climate Research

Advisory and Transformation Services | GreenLake Flexible Capacity | Datacenter Care Services

Supercomputing / Enterprise / Commercial HPC

HPE SGI 8600
HPE Apollo 6000 Gen10
HPE Apollo 2000 Gen10
HPE Apollo 70 ARM

HPC Software Portfolio

HPE Performance Software Suite
- HPE Insight Cluster Management Utility
- HPE SGI Management Suite
- HPE Performance Software – Cluster Manager (June 2018)
- HPE Performance Software – Message Passing Interface

HPE Storage Software
- HPE Data Management Framework

Emerging HPC

HPE Apollo 6500 Gen10

In-memory HPC

HPE Superdome Flex Server
Scale-up, shared memory HPC, combines best of HPE and SGI technologies

HPC Storage

HPE Apollo 4510 Gen10
HPE StoreEver LTO-8 tape

Networking

Intel® Omni-Path Architecture
Mellanox InfiniBand
HPE FlexFabric Network
Superdome Flex modular scale-up architecture
HPE Superdome Flex Technology
8th Generation HPE ccNUMA Memory Architecture

HPE Superdome Flex ASICs
- Scale beyond capabilities of Intel UPI links

Global memory
- All memory is shared by all processors

Hardware cache coherency
- Copy consistency at ASIC speed and directory-based

High bandwidth – low latency
- 13.3 GB/second; max read latency <400ns!
Superdome Flex Chassis – HPE Fabric

**HPE Fabric**
- Each chassis has two HPE Superdome Flex ASICs that connect the processors into the HPE fabric
- Each ASIC exposes 16 four-lane fabric links
- ASICs in each chassis are connected, leaving 15 links per ASIC for cabling to ASICs in other chassis

30 HPE Fabric Cable Ports
- Infiniband (EDR) electrical interface and cabling (not an Infiniband network)
- Fault tolerant if cable removed
HPE Superdome Flex

All-to-All between all HPE Superdome Flex ASICs for an 8 chassis system

- Scale single server from 4–32 sockets
- All-to-All topology
- Max 1 hop
  - 100 - <400ns
Modular, Future-Ready Design

Superdome Flex Scales up Seamless as a Single System

- Never outgrow
- No over-provisioning
- No forklift upgrades!

4-Socket (6TB)  8-Socket (12TB)  16-Socket (24TB)  20-Socket (30TB)  32-Socket (48TB)

12-socket, 24-socket, and 28-socket configurations not pictured
HPE Superdome Flex
Price/Performance Advantage over Smaller Systems

Superdome Flex can deliver a lower cost, higher performance 6TB solution than competitive 4-socket solutions

- Competitive 4-socket systems are forced to use expensive ‘M’ processors and 128GB DIMMs
- 8-socket Superdome Flex using 64GB DIMMs delivers double the compute power, double the memory bandwidth, and double the IO capability

Superdome Flex can deliver a lower cost, higher performance 8-socket solution than competitive 8-socket solutions

- Competitive 8-socket systems are forced to use expensive Platinum processors
- 8-socket Superdome Flex can use lower cost Gold processors

Superdome Flex can scale beyond 8 sockets!

- Scale up with lower cost Gold processors and 64GB DIMMs

Designed with Memory-Driven Computing principles
Superdome Flex Chassis Rear View

Base chassis pictured

- VGA port
- Serial port
- PCI Express slots 16 LP or 8FH/4LP

- Internal 4 x 2.5” HDD/SSD drives
- 1GbE RMC port
- 1GbE Management (KVM) port
- Four 1600W Hot-swap Power Supplies N+1 or N+N redundancy
- Hot-swap Fans for cooling HARP ASICs

Base I/O*
- DVD Drive
- 2 x 1GbE
- 2 x 10GbE
- 4 x USB 3.0

* For Expansion chassis, the Base I/O area is replaced with single RMC, Mgmt, and BMC ports
Mission-critical RAS
**End-to-end RAS delivers five nines system availability**

HPE Superdome Flex RAS features at a glance

<table>
<thead>
<tr>
<th>Chassis-Level</th>
<th>HPE Superdome Flex: key areas of RAS superiority</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>RAS</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Superdome Flex</strong></td>
</tr>
<tr>
<td>Firmware-first</td>
<td>✓</td>
</tr>
<tr>
<td>Automatic error logging</td>
<td>✓</td>
</tr>
<tr>
<td>Auto self-healing (Analysis Engine)</td>
<td>✓</td>
</tr>
<tr>
<td>Disabling / deconfiguration of failed FRUs</td>
<td>✓</td>
</tr>
<tr>
<td>Onboard fault analyzer</td>
<td>✓</td>
</tr>
<tr>
<td>Automatic restart</td>
<td>✓</td>
</tr>
<tr>
<td>Advanced processor error handling (eMCA Gen2)</td>
<td>✓</td>
</tr>
<tr>
<td>Advanced memory resiliency (ADDDC)</td>
<td>✓</td>
</tr>
<tr>
<td>Enhanced fabric resiliency (Adaptive routing)</td>
<td>✓</td>
</tr>
<tr>
<td>Advanced PCIe error recovery (LER)</td>
<td>✓</td>
</tr>
<tr>
<td>Hard Partitions (nPar)</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Processor**

- Enhanced I/O
- Integer pipelines
- ECC coverage
- Register/Tag compression
- Improved direct memory access
- UPI link-level recovery
- UPI rolling recovery
- Core level recovery
- Poison Data Recovery

**Architecture**

- Hyperthreading
- Dynamic socket migration
- Chip multiprocessor technology
- Improved direct memory access
- UPI link-level recovery
- UPI rolling recovery
- Core level recovery
- Poison Data Recovery
System Management
## HPE Superdome Flex server management features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Server management benefits from significant engineering investment</strong></td>
<td>New firmware (FW) provides Redfish API’s yielding OpenStack compliance, OneView*, and enterprise remote support</td>
</tr>
<tr>
<td><strong>RMC User Interface + new embedded RMC (eRMC) for 4s and 8s systems</strong></td>
<td>A comprehensive Linux CLI with Help features to guide users through command syntax and simplify tasks</td>
</tr>
<tr>
<td><strong>Advanced capabilities help diagnose and recover from faults</strong></td>
<td>Enhanced Error Analysis capability improves self-diagnosis (with future automated nPar recovery)</td>
</tr>
<tr>
<td><strong>Dramatically improved security for mission-critical</strong></td>
<td>Support of multiple unique admin/root account passwords. FW update will use signed/certified FW bundles (no more individual .rpm files to match and track)</td>
</tr>
</tbody>
</table>

---

* current-tech preview; CQ2-monitoring; CQ4- management
Storage
Superdome Flex supported storage

- **Internal storage:** Four (4) 2.5” drive bays to support SATA SSDs or SAS HDDs/SSDs.
  - 6G SATA SSDs use embedded chip (Intel RSTe) with SW RAID (w/boot support)
  - 12G SAS HDDs/SSDs use PCIe RAID card (internal) with HW RAID (w/boot support)
- **SAS:** HPE SAS JBOD (e.g. D3700) are supported with PCIe RAID card (external) (w/boot support)
- **Fibre Channel:** HPE FC arrays (e.g. 3PAR, XP, MSA) are supported with PCIe FC HBAs (w/boot support)
- **Third party storage:** Storage vendor takes the lead in documenting interoperability
- **NVMe:** R-Pool (Part of ISSP/HP XFS); currently RAID-10
- **HPC Cluster:** Infiniband, Omnipath, Ethernet (NFS); Multi-rail Lustre, HP XFS
The NVIDIA advantage!
Post processing visualization
HPC Visualization, workload acceleration, and AI

HPE Superdome Flex +

✅ NVIDIA® Tesla® P100
✅ NVIDIA® Tesla® V100
✅ NVIDIA Quadro® P6000
HPE In-Memory HPC solutions accelerate time to discovery

HPE
Superdome Flex

Solve complex, data-intensive problems holistically at greater scale

Complete more jobs in less time

Free HPC teams from managing clusters to accelerate time to discovery
- No code changes needed
- Like a giant workstation with lightning speed!
Questions?
Thank you