Supercomputer Best Practices Seminar

HPC Platforms in the Industry Solutions and Proof of Concept Centers

David V. Gelardi
VP, Industry Solutions and Proof of Concept Centers
dvg1@us.ibm.com

May 11/12, 2005
Agenda

- **Industry Solutions and Proof of Concept Centers**
  - Missions
  - Locations
  - Skills and Resources

- **Supercomputing center best practices**
  - Performance Tuning
  - Leveraging Results
Industry Solutions and Proof of Concept Centers

- **Design Center for on Demand Business**
  - First of a Kind for On Demand Technology

- **Customer Benchmark Proof of Concept**
  - Tactical, Closing Sales

- **High Availability Center of Competency**
  - Driving Value, Influencing Clients

- **Deep Computing Capacity on Demand**
  - P&L CPU’s by the Hour

**2004**
- 1,000+ Engagements
- $1B STG Revenue
- 7,000+ CPU’s
- 1.5 Petabytes Storage
- 15 Locations WW

**HW/SW Pool**
- WebSphere
- Blue Gene
- GRID
- DB2
- z/OS
- IBM TotalStorage
- BladeCenter
- iSeries
- pSeries
- SAP

**Skills Pool**
- IT Architects
- z/OS
- Virtualization
- AIX
- i5/OS
- HPC
- DB2
- Middleware
- Tivoli
- Cisco
- i5/OS
- OpenPower
- zSeries
- Security
- Linux
- Middleware
- Project Managers
- WAS
- Networking
- IT Specialists
- Oracle
- Windows
- System Administration

**STG Performance Marketing**
- eServer and TotalStorage
- Benchmark Planning & Marketing

**Development Center for Solution Integration**
- Solution Offering
- Architecture and Integration
HPC Missions

WW Customer Benchmark Centers

- Provide customer demanded benchmark capability WW for eServer and TotalStorage
  - Include proof of concept, scaling and performance
  - Assist in the execution of ISV Application benchmarks

Deep Computing Capacity on Demand Centers

- Develop and deploy a Capacity on Demand offering that provides clients with an alternative means to meet their peak computational needs, do so in a manner that is complementary with traditional server and storage sales.
  - Provide increased Business Value to our customers through innovative On-Demand delivery mechanisms.
  - Provide a competitive differentiator for IBM.
  - Drive incremental revenue and profit.

Performance Marketing

- Lead series in the creation of plans for industry benchmarks and assist in the marketing of results (e.g. TOP500)
  - Participate in planning with Series product marketing and development to optimize cross-series plans and leverage cross-series opportunities
  - Create eServer and TotalStorage marketing collateral and provide customized assistance to sales opportunities
What is a Client Benchmark

- **A Demand to Prove a 'Capability' to a Specific Client**
  - Often Part of an RFP Response
  - Often Competitive
  - Client Sometimes Needs Help Defining Requirements
  - Categories of Information are Performance, Scaling and Proof of Capability, etc.
  - Client Sets Criteria against their Data and Workload
  - Strict Response Date, Usually Very Short Term

- **Different from...**
  - Industry Standard Benchmarks (usually done in development)
  - TPC-C, TPC-H, SPECnnn, STREAM, Pallas, NPB, Linpack HPL etc.
  - Application Benchmarks (usually done in ISV Enablement)
  - SAP SD, Peoplesoft, BAAN, Siebel, etc.
The Challenge: Optimizing HPC Capacity Utilization

- Traditional infrastructure build-out increases in step-function phases
- Companies that build for average demand must be able to respond quickly to peak workload demands or suffer lost opportunity
- Companies that over-build to address peak workloads are left with over-capacity and under-utilization in business downturn
- IBM Deep Computing Capacity on Demand serves unfulfilled peak workload requirements

In-house Resources or Outsourced Data Center

- Unfulfilled Business Peaks
- Unutilized Resources
Management of Resources

- Multiple clients access “virtual” clusters made up of compute and storage resources
  - Dedicated to one client at a time
    - dedicated custom environment or
    - timesliced optimized environment (no root access, job scheduling)
  - Some resources open to all for testing, prototyping, tuning
  - Allocations subject to availability

- Highly secure and resilient infrastructure

- Client components
  - Application software and licenses
  - Data
  - Custom hardware
## HPC Benchmark Systems

<table>
<thead>
<tr>
<th>p5 Machines</th>
<th>Open Power</th>
<th>Clusters</th>
<th>Storage</th>
<th>Pentium4</th>
<th>Intel EM64T</th>
<th>Opteron250</th>
<th>Xeon</th>
</tr>
</thead>
<tbody>
<tr>
<td>p520 x 1</td>
<td>p720 x 4</td>
<td>32 x p575 w/ High Performance Switch</td>
<td>ESS x 1</td>
<td>x335 x 16</td>
<td>x336 x 36 w/ Myrinet and IB</td>
<td>e325 x 64</td>
<td>x345 x 26</td>
</tr>
<tr>
<td>p550 x 2</td>
<td></td>
<td>8 x p575 w/ Gigabit Ethernet</td>
<td>DS4500 x 26</td>
<td></td>
<td></td>
<td>e326 x 68 w/ Myrinet and IB</td>
<td></td>
</tr>
<tr>
<td>p570 x 26</td>
<td></td>
<td>16 x p710 w/ Gigabit Ethernet/Myrinet</td>
<td>SAN Switches x 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p595 x 3</td>
<td></td>
<td>2 x Blue Gene</td>
<td>EXP 700 x 117</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>p4 Machines</th>
<th>PowerPC</th>
<th>Pentium4</th>
<th>Intel EM64T</th>
<th>Opteron250</th>
<th>Xeon</th>
</tr>
</thead>
<tbody>
<tr>
<td>p615 x 1</td>
<td>JS20 x 8</td>
<td>x335 x 16</td>
<td></td>
<td>e325 x 64</td>
<td></td>
</tr>
<tr>
<td>p630 x 2</td>
<td></td>
<td></td>
<td></td>
<td>e326 x 68 w/ Myrinet and IB</td>
<td></td>
</tr>
<tr>
<td>p630+ x 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p655+ x 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p690 x 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p690+ x 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p690++ x 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Quantities can vary based on active projects*

*Access to Loaner equipment leveraged*
IBM DCCoD: Scale Beyond In-House HPC Limits

Secure Internet access to supercomputing power owned and hosted by IBM enables clients to rapidly and temporarily flex up/down HPC capacity proportional to business demands - to respond to peak workloads and capture business opportunities that would otherwise be out of reach.

= Fixed Capacity / Fixed Cost

= Variable Capacity / Variable Cost

4 DCCoD centers
- Intel® Xeon™
- AMD Opteron™
- IBM POWER™
- IBM Blue Gene®
- 7000+ CPUs

Poughkeepsie, NY
Montpellier, France
Rochester, MN

IBM HPC Grid

Virtual Private Network
IBM Blue Gene® Capacity on Demand

Flexible and convenient pay-for-use access to reserved capacity for scientists, researchers, and developers - to a new family of supercomputers optimized for scalability, bandwidth, and massive data handling while consuming a fraction of the power and floor space required by today’s fastest systems.

- World’s fastest supercomputer!
- Ultra scalable performance
- Ultra floor space density
- Ultra performance per W of power
- Innovative architecture and system design
- Familiar programmer/user Linux-based Environments

Multi-use on demand access to …
- Up to 2x 1,024 dual PowerPC® processor compute nodes per rack
- 64 IO Nodes per rack
- Front-End Node
- I/O File Server
- Service Node
  … on a sub-rack basis

Available March 2005

World's fastest supercomputer!

World’s fastest supercomputer!

Ultra scalable performance

Ultra floor space density

Ultra performance per W of power

Innovative architecture and system design

Familiar programmer/user Linux-based Environments

#1 on TOP500

70.72 TFLOPs sustained

Rack

2,048 PowerPC® CPUs

2/8/5.6 TF

512GB Memory
Skills

- **Scientific and technical application skills**
  - Architecture porting
  - Algorithms
  - Compilers: Fortran 77, Fortran 90, C, C++, Libraries & Tools
  - Application tuning
  - Parallel Programming, message passing
  - Performance analysis

- **System Administration skills**
  - Job scheduling: Load Leveler, PBS
  - OS skills: AIX, Red Hat, SLES distributions
  - System tuning
  - SW installs and systems management, ID management
  - Interconnect: High Performance Switch, Myrinet, Topspin IB, Voltaire IB
  - Disk subsystems, GPFS
  - Troubleshooting

- **Infrastructure Skills**
  - Hardware provisioning
  - Networking
  - Security and Access

- **Project Management skills**
HPC Typical Workloads

- **Universities**
  - **CAE:** Nastran, Star-CD
  - **Petroleum:** Reservoir Simulation
  - **Life Sci:** Gaussian
- **Kernels**
- **OpenMP Codes**
- **Cluster Size**
  - Standalone
- **Large**
- **High**
- **CAE:** Fluent, Abaqus
- **Life Sci:** Charmm, Amber
- **Petroleum:** Seismic
- **Gov't:** RYO
- **Weather:** MM5
- **Life Sci:** Blast
Performance Tuning Process

- Result correctness is the basis of all tuning
  - Reference data from customer
  - Comparison of results from very different levels of compiler optimizations.
- Platform analysis
- Profiling and analyzing code for potential improvements
  - Compiler optimization
  - Performance libraries such as IBM ESSL and MASS
  - Explore auto-parallelization by compiler and ESSLSMP (part of ESSL library) library
  - Apply hand tuning (optimization, OpenMP parallelization and/or MPI parallelization) based on the profile
- Repeat, Repeat, Repeat
  - Within time and cost/benefit constraints
Discussion Points

- Mission – What is it?
- Planning is Key
- Infrastructure is Key
- Security a Nightmare
- Power / Cooling
- Skills
- Broad Applications (and long lived!)
- No Test Systems (FOAK)
- SLAs
- Chiba City Like OS/Software Stack
- Benchmarking v. Production (DCCoD)
- Metrics / Cust Sat / Surveys
- User Interaction
- Rapid Turnover of Assets In/Out
- What is not 7x24x365?