Virtualization + Grid Computing = Data Center Transformation

John Williams
Director
Server and Workstation Division
AMD
Optimizing the Data Center

- Replace older hardware with more operationally efficient servers
- Use virtualization to consolidate multiple workloads to improve utilization
- Buy power efficient dual-core servers to help decrease the power, cooling, and space requirements
- Maintain compatibility with existing software infrastructure
- Decrease manageability and maintenance requirements
# Transforming the IT Infrastructure

## Dedicated Infrastructure
- Infrastructure funded by projects
- Significant unused capacity
- Inflexible
- Many legacy technologies

## Shared Infrastructure
- **Virtual Resource Pool**
- Change how infrastructure is funded and planned

### Benefits
1. **“One infrastructure to build it all”**
   *Extreme Modularity and Integration*
2. **“Any workload, just in time.”**
   *Pervasive Virtualization and Automation*
3. **“Managed as one”**
   *Unified Server, Storage, Network Control*
Key Innovations Driving Transformation

64-bit and multi-core processors are enabling smaller, denser servers and blades with high-end performance.
Virtualization Technology

Breaks the “One Application to One Server” Model that Is Driving Under Utilization of x86 Servers

Single application “owns” all available processors, even if it does not use them

Single application uses some of available processors
**X86 Virtualization Evolution**

<table>
<thead>
<tr>
<th>x86 Virtualization</th>
<th>Software Virtualization</th>
<th>Silicon Assisted Virtualization</th>
<th>Device Virtualization</th>
<th>Performance Tuning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMD64 Technology</td>
<td>Direct Connect Architecture</td>
<td>Virtualization Extensions</td>
<td>I/O Virtualization</td>
<td>Nested Page Tables</td>
</tr>
</tbody>
</table>

- **Direct Connect Architecture**: Increase performance with multi-core capabilities and fast memory access.
- **Virtualization Extensions**: Simplify virtualization software and reduces CPU overhead.
- **I/O Virtualization**: Improve performance and security of virtualized devices.
- **Nested Page Tables**: Reduce the overhead of switching between machines.
What is Grid Computing?

- Configuration of industry-standard 2P and/or 4P servers

- Creates a virtualized set of shared services that can be dynamically allocated to meet demands

- Reduces space, power, hardware, and administration costs while providing an efficient and secure computing infrastructure
Grid Computing - the New Architecture for Business Solutions

- Non-proprietary
  - Cluster of industry-standard servers
  - Uses any operating system

- Consolidates resources
  - Creates a single set of shared services
  - Resources allocated as needed

- Performance and scalability
  - Leverages 64-bit and dual-core technology
  - Dynamically provisions resources

- Efficient and secure infrastructure
  - Centralized management
  - Adjusts resources without delay
The AMD Advantage in an Optimal Platform
Driving Innovations into the x86 Processor

64-bit and Dual-Core

Enhances performance while offering the flexibility to support both 32- and 64-bit applications

Performance-per-watt

Assists data centers in controlling power consumption and heat output

Direct Connect Architecture

Eliminates the 20-year old front-side bus, increasing system efficiency and scalability

Virtualization

Increases utilization by enabling the running of separate, secure operating environments
Grid Computing & Virtualization: Creating the Optimal Platform

- **Compute Density/Scalability**
  - Chip and platform level
  - Increased utilization with virtualization

- **Power/Cooling**
  - Performance/Watt is as important as raw performance
    - Varying definitions of ‘performance’ by workload
    - Raising the numerator is as important as lowering the denominator
  - Increased performance with each additional Watt
    - “No Watt Left Behind”

- **Workload Affinity**
  - Single thread performance still matters – Amdahl’s Law
  - Migration to multi-core: new challenges/opportunities
  - Complex, multi-threaded virtualized environments
Grid Computing & Virtualization

- Compute Density/Scalability
- Power/Cooling
- Workload Affinity
Grid Computing & Virtualization: Maximizing Compute Density

- Within the Microprocessor...
  - Multi-Core solutions

- Within the Server...
  - Direct Connect
    - Efficient, linear scalability
    - More performance within a given form factor
  - AMD Virtualization Technology

- Within the Data Center...
  - AMD PowerNow!
    - Increased rack utilization
  - AMD Virtualization Technology
    - Server consolidation and management
    - Secure virtual environments
Within the Microprocessor...
Why Multi-Core?

- Single core is ultimately limited by thermal density
  - Example: Migration from 130nm to 90nm
    - 50% core area reduction – same total power = \(2x\) Watts/mm\(^2\)
    - Power is concentrated at the CPU core

- Multi-Core is the evolution of Moore’s Law
Within the Microprocessor...
Multi-Core: Power Efficiency & Performance

Quad Core Processors:
~15% Freq. Reduction vs. Dual
50-70% Performance Increase
Same Power Consumption

Approximate limit for air-cooling

Value of “N” continually increases with transistor technology improvements

*Based on 2.8GHz Opteron™ vs. 2.6 GHz Opteron HE
Within the Server...

Scalability: AMD Direct Connect Architecture

- **It’s about the System**
  - Balanced architecture
  - Flexible architecture
  - Linear scaling to 4P and beyond

- **Compute Scalability**
  - CPU to CPU connectivity
  - No limiters for multi-core

- **Memory Scalability**
  - Minimal latency
  - Add capacity with processors
  - Add bandwidth with processors

- **I/O Subsystem Scalability**
  - Add bandwidth with processors

---

Every AMD Opteron™ Processor supports it’s own memory & I/O
Within the Data Center...
AMD Virtualization Technology

- Server Virtualization and Consolidation
  - Multiple operating environments on a single platform
    *Isolation of independent virtual machines*
  - Higher Server utilization/less hardware
    *Net performance/Watt gain*
  - Reduced hardware management costs

- Server Consolidation – what’s the catch?
  - High peak loads
  - Potential hidden costs
    *Software license costs on larger 4P or 8P versus 2P*
    *Diminishing returns on scale-up beyond 4P or 8P*
    *Still have the same (or greater) number of OS’s to manage*
  - A great opportunity – not the answer to everything
Evolution of Enterprise Computing

- Compute Density/Scalability
- Power/Cooling
- Workload Affinity
Managing Power & Cooling

• Maximizing Performance per Watt
  – Multi-core processor solutions
  – Microarchitectural enhancements
  – Next generation manufacturing technologies

• Power-efficient Server processor solutions

• Advanced power management
  – AMD Power Now! – scaling power consumption with demand

• Optimal system scalability
  – Getting the most performance out of every Server
  – Keeping memory latency low
  – Keeping memory and I/O bandwidth high
Evolution of Enterprise Computing

- Compute Density/Scalability
- Power/Cooling
- Workload Affinity
Workload Affinity:
Multi-Core Evolution: One Size Does Not Fit All

- Workload Variations
  - Database, Transaction Processing, Search
    Multiple unrelated tasks/sessions supported at one time
    More threads are good (inherently parallel)
    Integer intensive
    Moderate single thread performance needs
  - Technical Computing, HPC, Workstation
    One problem broken into multiple tasks
    Larger data sets
    Less threads
    Floating point intensive
    High single thread performance needs

- Single core history
  - Increasingly higher single thread performance, limited threading
  - Good for Technical Computing/HPC/Workstation
  - Ok for Database/Transaction Processing coupled with SMP
Workload Affinity: Extending Performance/Watt with Accelerators

- Heterogeneous multiprocessing
  - Extending solutions beyond basic core/frequency combinations
  - General purpose AMD64 x86 capabilities
  - Specialized engines for specific workloads

- Specialized engines maximize performance/Watt
  - Example: HD video decode – 300 MHz 32-bit MIPS CPU

- Acceleration engine opportunities
  - Vector/Floating Point
  - Security
  - Media Processing
  - Managed Code (Java/XML)
  - Others...
Driving x86-based Virtualization

AMD is developing a robust virtualization ecosystem that provides users with innovation and choice.
AMD64 + Red Hat Linux
Solving Today’s IT Challenges

• Multi-core processors
• Server virtualization
• Grid computing

• Accelerate application performance
• Improve asset utilization
• Reduce computing complexity
• Control infrastructure costs
Thank You
Trademark Attribution

©2006 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, and combinations thereof, are trademarks of Advanced Micro Devices, Inc. Other names are for informational purposes only and may be trademarks of their respective owners.