# Solaris for High Performance Computing

**Matthew Baier**

Product Line Manager, HPC

Solaris Marketing

Sun Microsystems, Inc.

# Agenda

**What's New**

The Importance Of Being Open

Better Than Insulin (And Other Cool Stuff)

Performance

Industry Momentum

# What's New?

IBM jumps on the Solaris bandwagon

vnunet.com

Solaris, Java Go Open Source

The New York Times

Cluster File Systems, Sun to Integrate Lustre with ZFS

HPCwire

Ian Murdock: What's a Linux Guy Doing at Sun?

InfoWorld

AMD/Sun Strategic Alliance; Opteron to Power Sun's Servers

eWEEK THE ENTERPRISE NEWSWEEKLY

Sun Moves to Number Six x86 Server Vendor in 2005

IDC Analyze the Future

IDC Worldwide Quarterly Server Tracker November, 2006

Solaris Technology Wins WSJ's 2006 Top Innovation Award, Beating Inhalable Insulin

THE WALL STREET JOURNAL.



Sun, Intel Landmark Alliance; Intel Endorses Solaris

The Mercury News
The Newspaper of Silicon Valley
MercuryNews.com

Univa Corporation Announces New Relationship with Sun

YAHOO! FINANCE

# Refreshed HPC Solaris Strategy

Powered by Solaris, Sun introduces an all-new, open petascale architecture, and provides the strongest and most complete solution for HPC computing.

Proven, unrivalled

- Availability
- Reliability
- Security

# Solaris HPC Software Stack

Sun CRS, Support, Architectural, Professional Services

| | | |
|---|---|---|
| **Developer Tools** | Sun Studio<br>Sun HPC Cluster Tools | Free |
| **Management**<br>Workload Management<br>Cluster Management | Sun Grid Engine Software<br>Sun Connection, ROCKS, Ganglia | Open, Free |
| **Operating System**<br>**and Cluster File System** | opensolaris<br>solaris<br>*Open Storage Server*<br>*ZFS, S-QFS*<br>*Lustre, p-NFS* | Open, Free |
| **Compute Node**<br>Processor | 64 Bit<br>AMD 64 Opteron<br>intel<br>ULTRASPARC | Open |
| **Interconnect** | IB Switch NEM for Blades<br>3456 Port Non-Blocking Switch | Open |

# Agenda

What's New

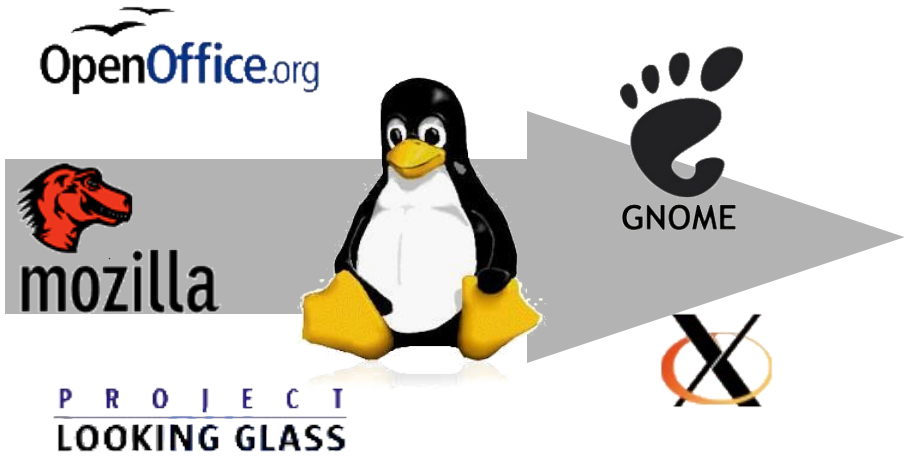
**The Importance Of Being Open**

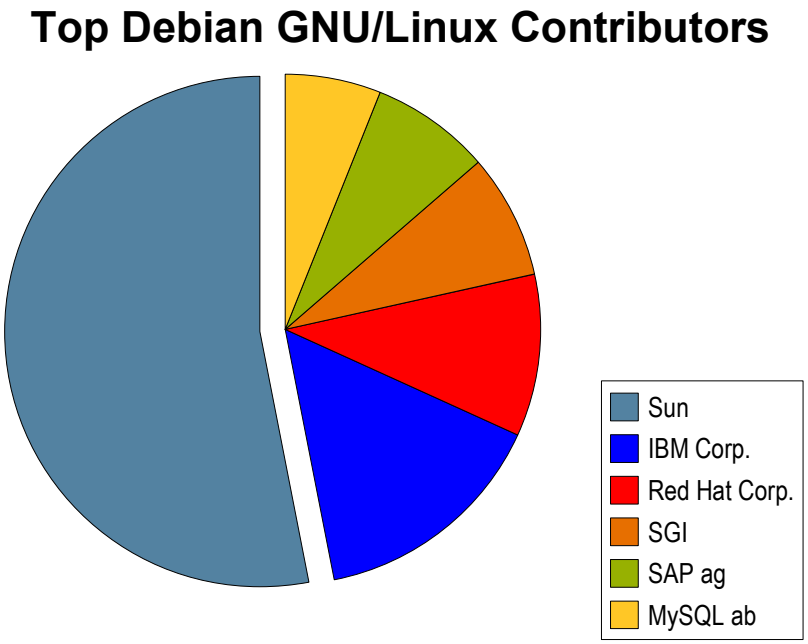
Better Than Insulin (And Other Cool Stuff)

Performance

Industry Momentum

# Sun Contributes to Open Source




## Solaris 10 incorporates 188 open source projects

Source: *Study on the: Economic impact of open source software on innovation and the competitiveness of the Information and Communication Technologies (ICT) sector in the EU*, November 20, 2006, UNU-MERIT for the European Commission, p. 51

opensolaris™

- Faster time-to-market for innovation
- Promote community participation
- Major worldwide adoption (Aug 07)
  - 5 distributions
  - 67,000 registered members
  - 42 user groups
  - 31 active projects
- SIIA Codie Best Open Source Solution award (May 2006)

# An Open Source Community

# Agenda

What's New

The Importance Of Being Open

**Better Than Insulin (And Other Cool Stuff)**

Performance

Industry Momentum

# Guaranteed Compatibility


Solaris 2.6 to Solaris 10

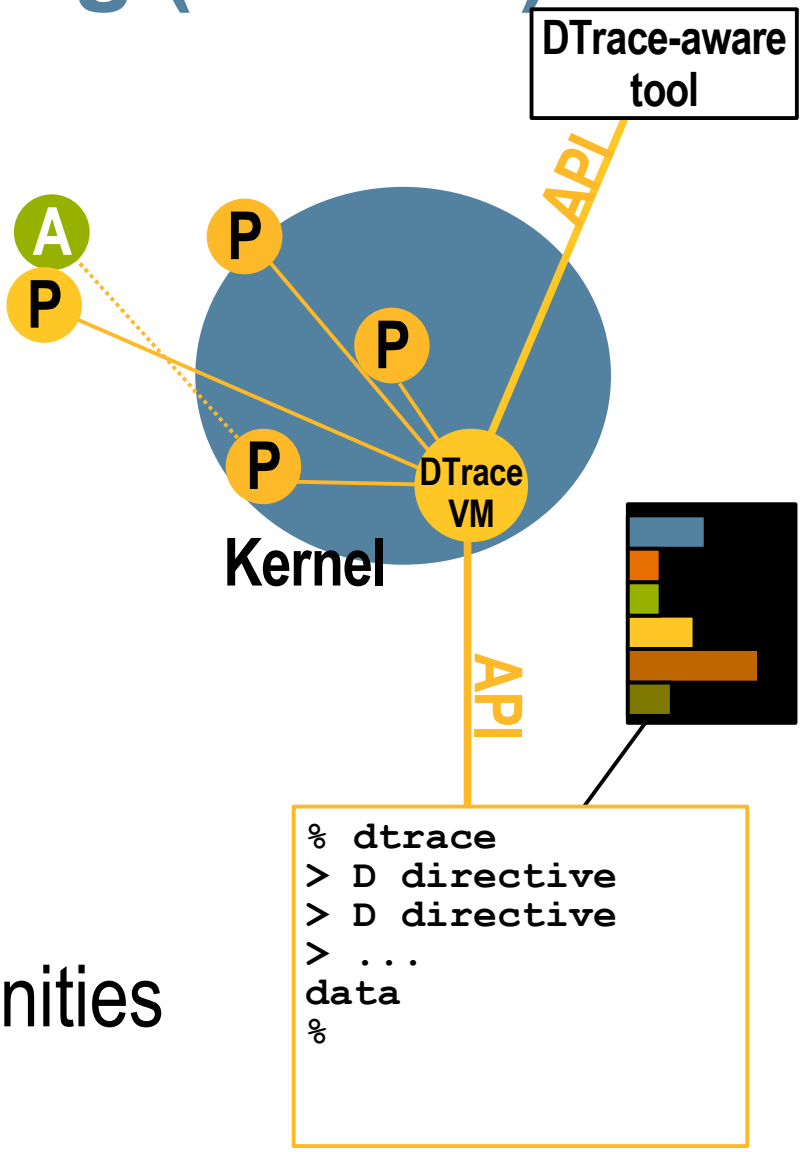
SPARC & x86

Lower Development & Support Costs

# Solaris Dynamic Tracing (DTrace)

## Designed for Production Systems

- Safe, always there
  - > No performance hit
  - > No app or OS changes
- Problems solved in minutes, not days
- Instrument every line in every application
- Views system as a whole
  - > Comprehensive
  - > Extensible, scriptable
- Massive performance opportunities



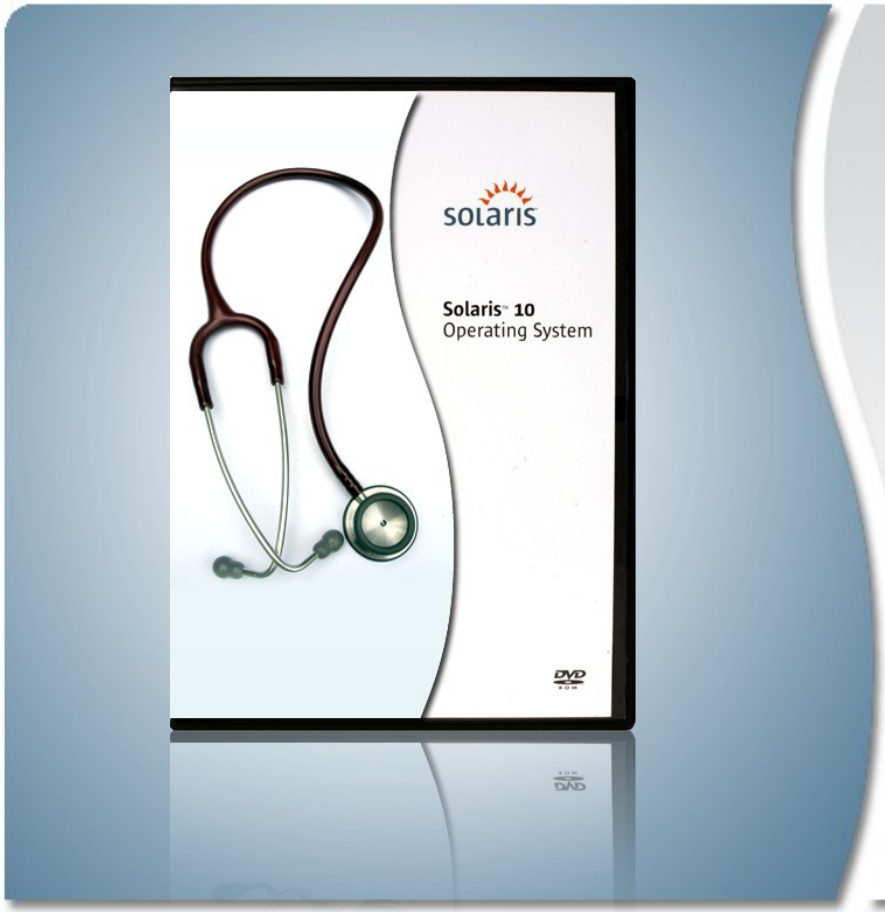
# Sun and InfiniBand


- High performance interconnect for HPC
- Sun an original driver of InfiniBand
  - Sun/Intel collaboration, driving the IB specification
  - A founder of OpenFabrics (formerly OpenIB)
- Complete IB infrastructure delivered in Solaris 10
  - IP over IB, uDAPL (user Direct Access Programming Library), SRP (SCSI RDMA Protocol), x86 boot over SRP etc.
- Support for all major IB providers
  - Mellanox, Voltaire, Cisco, QLogic (PathScale/Silverstorm)

# Predictive Self Healing

## Stops Problems Before They Happen


- Self-diagnosis in milliseconds
- Fine-grained recovery
- Zero source code changes required
- Close relationship with AMD, Intel, SPARC
- Solaris Fault Manager
- Solaris Service Manager




- **46%** reduction in annual downtime
  (6 CPU, 192 GB)
- **32%** reduction in annual downtime
  (4 CPU, 16 GB)
- **75%** faster boot time

# Superior Availability and Robustness

# Agenda

What's New

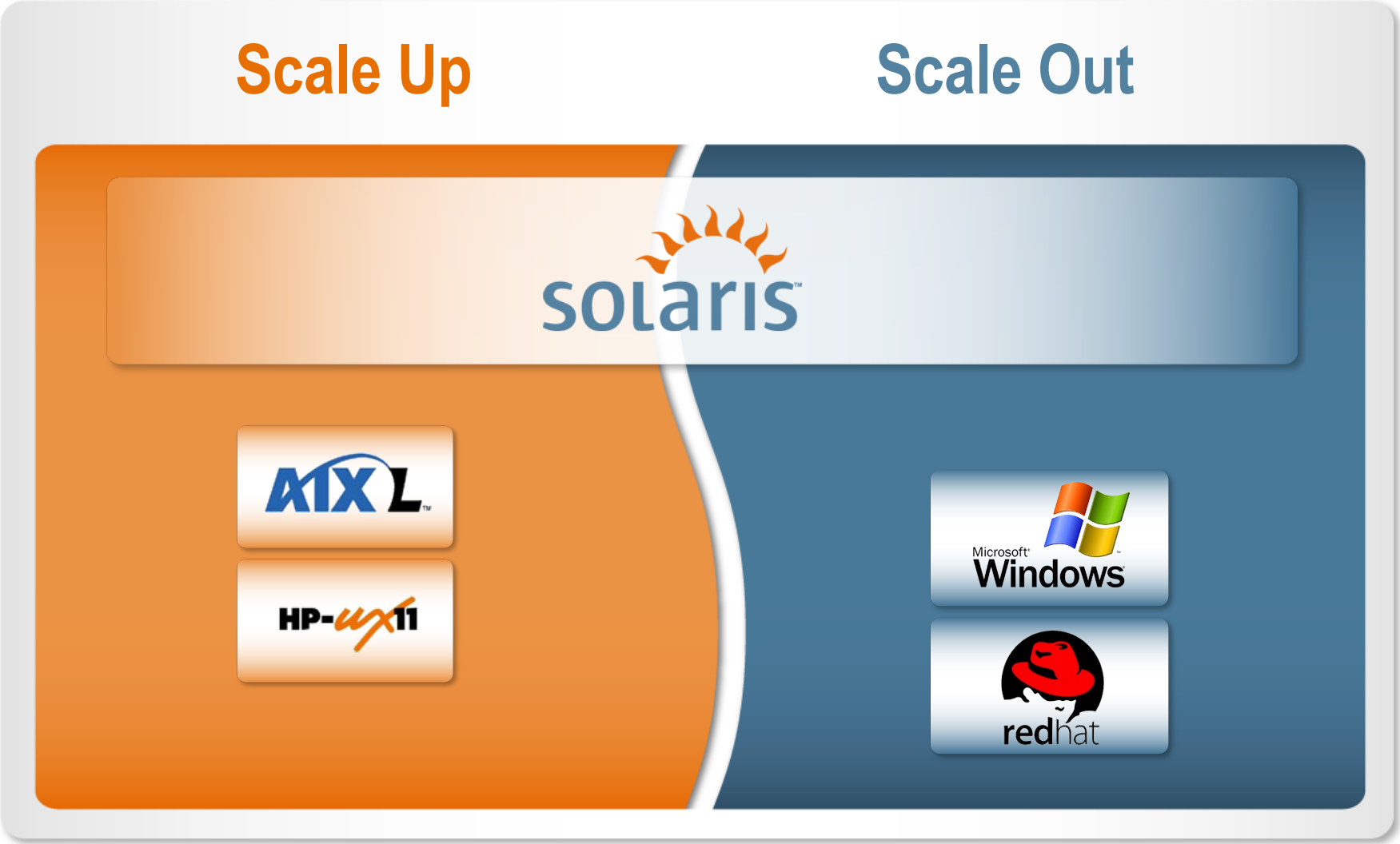
The Importance Of Being Open

Better Than Insulin (And Other Cool Stuff)

**Performance**

Industry Momentum

# Scales Across Your Infrastructure



# Solaris Scalability


- Vast expertise and investment in Multi-Threading
- 64-bit maturity
- Large CPU counts once the domain of large SMPs only
  - Fatter nodes now becoming the norm in clusters, with larger memory and I/O
  - Multi-threaded, multi-core CPUs
  - Integration of multiple boards onto a single die
- Solaris made this investment years ago
  - Proven scalability to 144 cores in a node, and soon beyond
  - Superior memory handling
  - Networking improvements

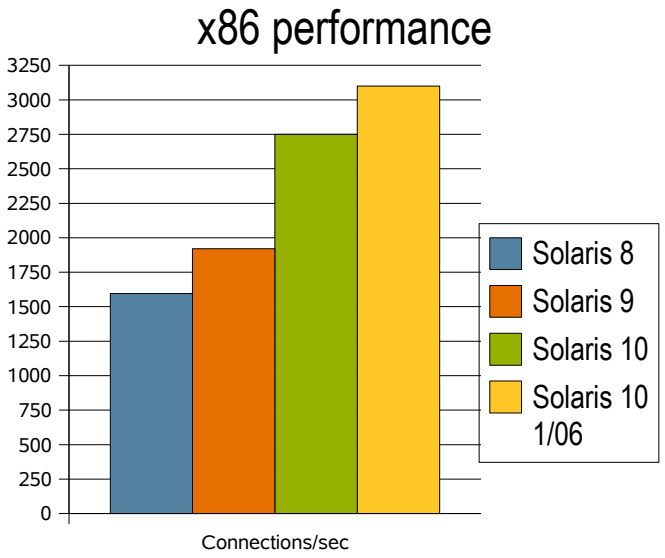
# Solaris HPC Performance Enhancements (1/2)

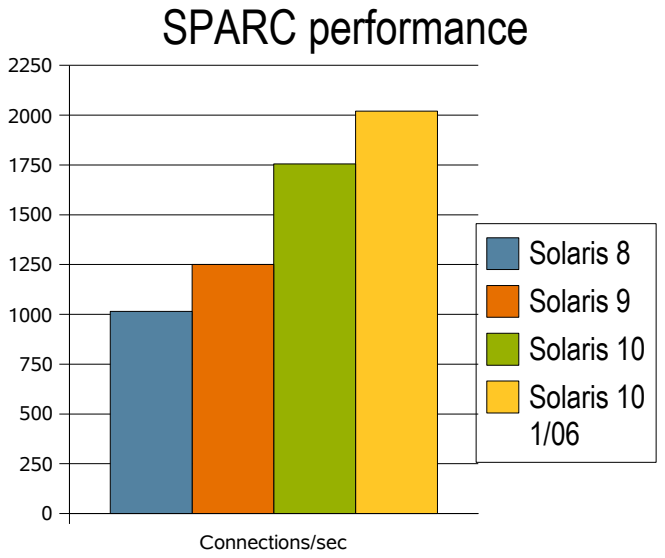
- Memory Placement Optimization (MPO)
  - Optimizes for performance through latency and bandwidth
  - Places memory as close as possible to the processors that access it while maintaining balance within the system to avoid bottlenecks
  - Delivers improved performance without requiring changes to most applications
- Hierarchical Latency Group Support (HLS)
  - Load balances across strands, cores, and/or system boards
- Multiple & large page sizes
  - Benefits performance-sensitive applications using large memory/data sets
  - Reduces Translation Lookaside Buffer (TLB) “misses”
- OpenMPI performance optimizations

# Solaris HPC Performance Enhancements (2/2)

- DTrace providers for many applications including Sun Grid Engine
  - Superior observability and performance analysis capabilities
  - Troubleshoot systemic problems in real time
  - Integration with OpenMPI stack brings DTrace benefits to parallel application development
- Dynamically-sized Translation Storage Buffers (TSB)
  - Efficiently utilizes hardware capability of supporting different TSB sizes
  - Avoids TLB “misses” and “cache-thrashing” between kernel TSB with user's TSB
  - Eliminates overhead of unnecessary TSB flushes
  - Eliminates the sharing of TSBs by multiple processes
- Large memory optimizations
  - Intimate Shared Memory – multi-process sharing of kernel translation structures

# Extreme Network Performance




- Performance has doubled since Solaris 8
- Still more to come...

# sun.com/solaris/benchmarks

8/24/06 Sun snatches two World Records in a brand new SPECcpu2006 benchmark
8/16/06 Lotus NotesBench R6Notes on Sun Fire V890
8/15/06, Sun Fire E25K LHS BSCS iX World Record Telco Billing
8/15/06, Sun Fire E25K SPECint_rate2000
08/15/2006 World Record 2-thread performance on SPEC OMPM2001 HPC benchmark for single socket systems
08/15/2006 The fastest single socket x86 system on floating-point throughput suite of SPEC CPU2000 benchmark
06/23/2006 The Sun Fire X4100 server delivers 100 GB TPC-H price/performance world record
06/23/2006 Best absolute price/performance and best in class performance on 300GB TPC-H benchmark
06/12/2006 1Sun's solution demonstrates the best performance per dollar on SPECjAppServer2004 benchmark
5/27/2006 Best performer on the new version of OCUS Benchmark
05/23/06, Sun Fire T1000 Server World Record SPECweb2005 Performance for Servers in 1U Footprint and World Record Performance per Watt
05/17/2006 Two world records: best 2-socket/2-way JVM scalability and throughput!
05/16/06, Sun Fire T1000 Server World Record SPECjbb2005 Performance for 1 Rack Unit / 1 Socket Servers and World Record Performance per Watt
05/15/06, Sun Fire T2000 Server World Record SPECjbb2005 Performance for 2 Rack Unit / Single Socket Servers
04/21/2006 MySQL database on Solaris 10 shines on OLTP Benchmark
Apr. 2006 Solaris 10, SAP TRBK, UltraSPARC E6900
Apr. 2006 Solaris 10 SPEC CPU2000 1s, AMD Sun Fire x2100
Apr. 2006 Solaris 10 SPEC CPU2000 2way, AMD Sun Fire x4100
Apr. 2006 Solaris 10 SPEC CPU2000 2way, AMD Sun Fire x4200
Apr. 2006 Solaris 10 SPEC CPU2000 workstation, AMD Sun Ultra 40
Apr. 2006 Solaris 10 SPEC CPU2000 throughput, AMD V40z
Apr. 2006 Solaris 10 Igen OLTP Database, UltraSPARC Sun Fire T2000
Mar. 2006 Solaris 10 SPECjAppServer2004, UltraSPARC Sun Fire T2000
Mar. 2006 Solaris 10 Floating point throughput dual socket, AMD Sun Ultra 40
Mar. 2006 Solaris 10 Lotus Domino NotesBench R6Notes, UltraSPARC Sun Fire T2000 US T1
Feb. 2006 Solaris 10 SPECjbb2005, UltraSPARC Sun Fire E25K US IV+
Jan. 2006 Pre-Installed with Solaris 10; Sun Announces Industry's Fastest x64 Workstation
Jan. 2006 The Sun Fire X4200 server delivers three-in-one punch on SPECjbb2005 benchmark

***World record floating point throughput on 4 socket x86 systems***

***Fastest x86 system for integer throughput calculations***

# 174 World Records

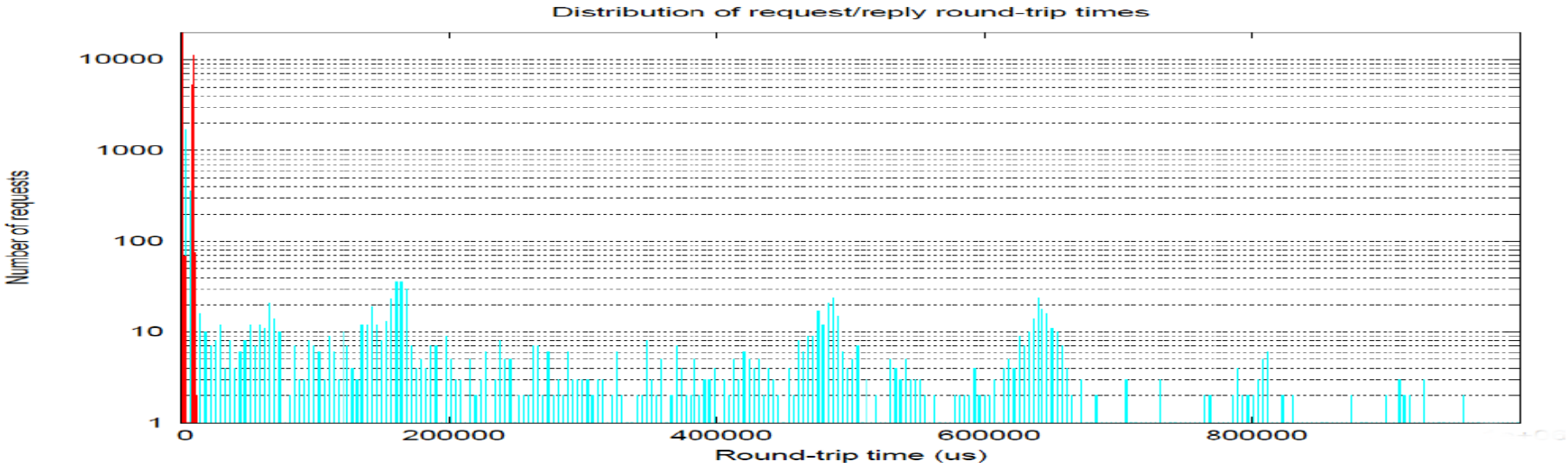
Nov. 2005 Sun's UltraSPARC IV+ Processor-Based Sun Fire Servers Continue to Beat IBM Power5 Systems
Oct. 2005 Sun Surpasses IBM for Server Leadership: New High-End Sun Fire Systems Give Sun New Edge; World-Record Benchmarks Released
Oct. 2005 Sun Studio 11, on the Solaris 10 Operating System, Delivers World Record Performance Results on Sun Fire Servers
Sep. 2005 Sun Fire X4100 Server - Benchmarks (five Solaris 10 benchmarks)
Sep. 2005 Sun Fire X4200 Server - Benchmarks (five Solaris 10 benchmarks)
Sep. 2005 Sep. 2005 Sun Microsystems Servers with Solaris Operating System and UltraSPARC Microprocessors Deliver Up to Fivefold Performance Boost
Sep. 2005 Sun Launches Blockbuster Family of Industry-Standard, x64 Enterprise-Class Systems (3 world records)
2005 Solaris 10 on Sun Fire E25K Achieved World Record Throughput on the Informatica Extract, Transform, and Load(ETL) Benchm
May 2005 Solaris 10 on Sun Fire E20K Set New Performance and Price Performance World Records on Manugistics Benchmark (2 world records)
May 2005 Solaris 10 on the Sun Fire V440 server running Sybase IQ set a new price/performance world record on TPC-H @300 GB
May 2005 Solaris 10 on the Sun Fire V240 server running Sybase IQ set a new price/performance world record on TPC-H @100 GB
May 2005 Solaris 10 on the Sun Fire V490 server running Sybase IQ set a new single-system price/performance world record on TPC-H @1000 GB
May 2005 Solaris 10 on the Sun Fire V880 server running Sybase IQ set a new single-system price/performance world record on TPC-H @1000GB
Solaris 10 on the Sun Fire E25K server running Oracle Database 10g set a new single-system price/performance world record on TPC-H @
2005 Solaris 10 and the Sun Java Application Platform Suite achieved best-in-class price/performance results on SPECjAppServer2004
Apr. 2005 Solaris 10 on the Sun Fire V40z server delivered record performance on eight-thread SPEC OMPM2001 HPC benchmark
Apr. 2005 Solaris 10 on the Sun Fire V40z server set a new world record on SPEC CPU2000
2005 Solaris 10 and the Sun Studio 10 software on the Sun Fire V40z server set new two-thread and four-thread world records on SPEC OMPM2001 (two world
Mar. 2005 Solaris 10 on the Sun Fire V40z server delivered world-record four-way 64-bit performance on SPECjbb2000
Mar. 2005 Solaris 10 on the Sun Fire V20z server at 2.6 GHz delivered world-record two-way 64-bit performance on SPECjbb2000
Feb. 2005 Solaris 10 on the Sun Fire E6900 server and BEA delivered world-record performance on SPECjAppServer2002 Dual Node
Feb. 2005 Solaris 10 on the two-way Sun Fire V20z server achieved a world record on SPEC OMPM2001
Feb. 2005 Solaris 10 on the Sun Fire E6900 server at 1.2 GHz delivered the batch Oracle Applications Standard Benchmark (HVOP) submission
10 on x64 delivered world-record one-way, two-way, and four-way 64-bit results on SPECjbb2000 on the Sun Fire V20z and Sun Fire V40z servers (three world records)
Jan. 2005 Solaris 10 on the Sun Fire E25K server delivered world-record performance on TPC-H @3000 GB
Dec. 2004 Sun Fire E4900 server with Solaris 10, the Oracle Database 10g, and the Sun StorEdge 6120 array delivered the first batch Oracle Applications Standard Benchmark (HVOP) submission
Nov. 2004 Solaris 10 on the Sun Fire V20z and Sun Fire V40z servers set a world-record aggregate bandwidth peak in the fifth-annual SCInet Bandwidth Challenge (two records)
Oct. 2004 Solaris 10 on the AMD Opteron processor-based Sun Java Workstations performed 34 percent to 61 percent better than comparable Intel processor-based Dell systems on Red Hat Linux 3.0 on BLAST (two records)

***Only company with published performance results on open source software***

***World record single system application server performance***

# Application “Jitter”

- Also called non-deterministic application latency
- For financial services applications, microsecond latency variations can be critical
- Even when general performance is within an acceptable range, latency can spike for various and sometimes unknown reasons
- Spikes of an order of magnitude can have significant business implications; Traders may have to abandon positions



# Solaris 10 – Curing Application Jitter

- Deterministic latency not easily achievable with operating systems not designed with “real-time” capabilities
- Preemptive kernel is critical to deterministic dispatch latency
- Solaris kernel is fully preemptable: Threads can interrupt another, lower-priority thread
- Observability using DTrace enables root cause analysis of issues underlying jitter
- Multiple scheduler support; low latency scheduling and interrupt dispatch
- Solaris provides the ability to lock memory to prevent paging/swapping
- Solaris provides for IP QoS through bandwidth manager
- The combination of the preemptive kernel, kernel interrupts as threads, fixed-priority scheduling, high-resolution timers, and fine-grained processor control makes Solaris an ideal environment for real-time applications or applications that require highly deterministic latency profiles
- Use case: **http://uadmin.blogspot.com/2006/05/what-is-dtrace.html**

# Agenda

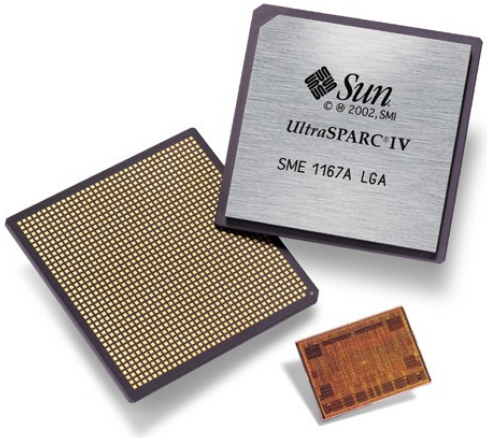
What's New

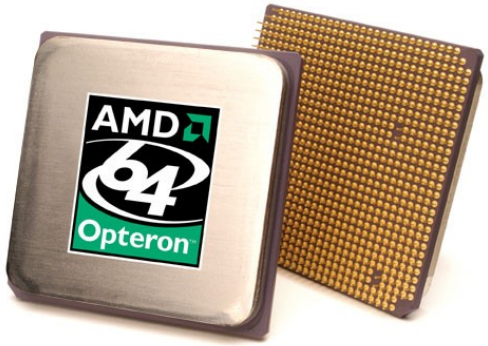
The Importance Of Being Open

Better Than Insulin (And Other Cool Stuff)

Performance

**Industry Momentum**





# Hundreds of Systems

Sun microsystems

IBM

DELL

hp

FUJITSU

TADPOLE

acer

SONY

egenera

TOSHIBA

lenovo

# Open Source Applications

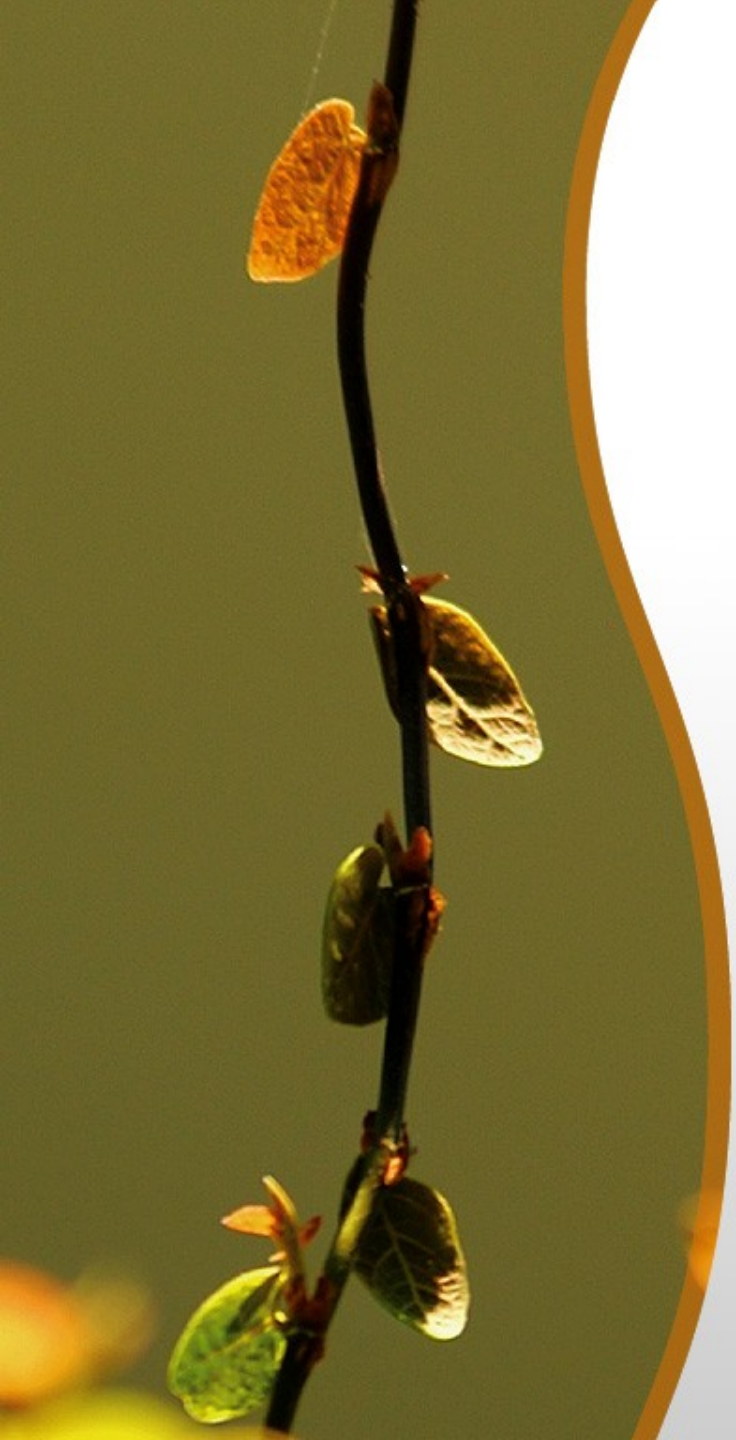
- HPC ecosystem and leading open source tools
  - ROCKS cluster provisioning for Solaris
  - Ganglia
  - Globus Toolkit
  - GD
  - BLACS
  - GPLK
  - HDF5
- See **http://www.sun.com/bigadmin/apps**

# Solaris – The Safe Choice

- Ubiquity
  - Broad platform support for non-Sun systems
  - IBM System X and BladeCenter systems
  - HP Blade servers
  - Available for x86/x64 and SPARC from Sun
- Used today in mission critical environments
  - Government and military
  - Stock market
  - ERP
- Investment Protection
  - Most cohesive stack – support from single vendor
  - Predictable roadmap with long lifecycle and support

# Solaris – Ideal For Commercial HPC

- A highly suited, proven platform for small to mid-sized clusters with unrivalled
  - **Availability**
  - **Reliability**
  - **Security**
- Addresses pain points in commercial HPC
  - Support for entire software stack from single vendor
  - Open, flexible architecture
  - Seamless scalability all the way to the high-end of HPC



# Summary

- World-class performance
- Maximum efficiency
- Open source and open architecture
- Call to Action
  - Visit **sun.com/hpc**
  - Visit **http://sun.com/solaris/move/**

# Resources

- HPC Learning Center:
  - http://sun.com/solaris/hpc (COMING SOON!)
- OpenSolaris HPC Community:
  - http://www.opensolaris.org/os/community/hpcdev
- HPC blogs:
  - http://blogs.sun.com/HPC
  - http://blogs.sun.com/marchamilton
  - http://blogs.sun.com/simons


Sun
microsystems