SPARC M6
Oracle's Next Generation Processor for Enterprise Systems

Ali Vahidsafa,
Senior Principal Hardware Engineer
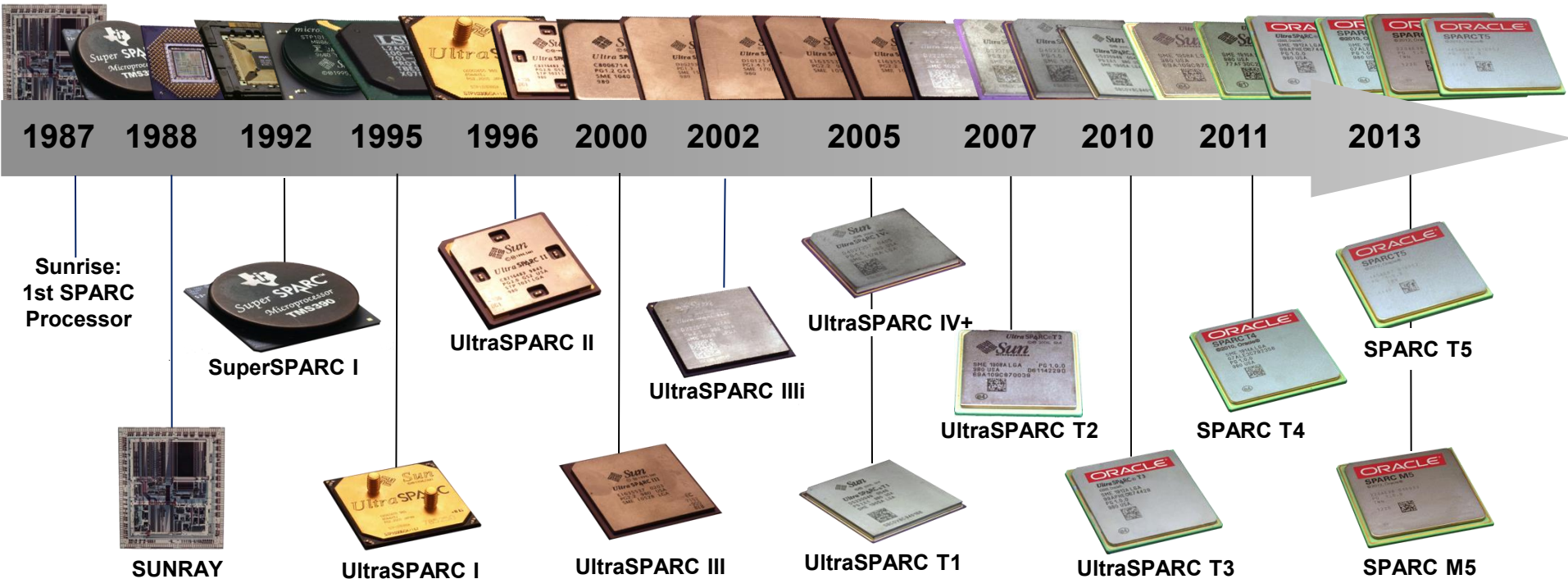
Sutikshan Bhutani,
Hardware Director

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

ORACLE

Outline

- Lineage
- Features
- Scaling
- Reliability, Availability, Serviceability
- Summary

ORACLE

25+ Years of SPARC Processors

Timeline: 1987, 1988, 1992, 1995, 1996, 2000, 2002, 2005, 2007, 2010, 2011, 2013

Sunrise: 1st SPARC Processor

SUNRAY

SuperSPARC I

UltraSPARC I

UltraSPARC II

UltraSPARC III

UltraSPARC IIIi

UltraSPARC T1

UltraSPARC IV+

UltraSPARC T2

UltraSPARC T3

SPARC T4

SPARC T5

SPARC M5

ORACLE

Dynamic Threading

- Hardware automatically re-allocates resources of inactive threads

- Results in multi-fold increase of single-thread performance

- Enables sophisticated OS scheduling - Solaris Critical Thread

- SPARC S3 Core achieves this without compromising compactness

- Extends benefits of massive-threading to a broader set of workloads

ORACLE

Current Products With Dynamic Threading

	nm	Cores	Threads	L3 Cache	Memory per Socket	PCIe	Max. Sockets
T4	40	8	64	4MB	0.5TB	2*G2	4
T5	28	16	128	8MB	0.5TB	2*G3	8
M5	28	6	48	48MB	1TB	2*G3	32

ORACLE

SPARC S3 Core & Oracle Solaris: Twenty #1's

- **Two #1s Database**
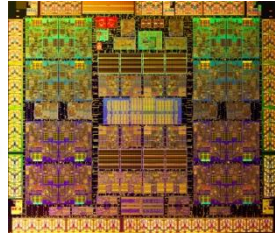 - #1 single-server TPC-C
  - #1 single-server TPC-H 3TB
- **Eight #1s in Applications**
  - #1 Java: SPECjEnterprise2010, #1 virt SPEC jEnterprise
  - #1 Java 2-chip: SPECjbb2013
  - #1 JD Edwards Online/Batch, #1 JD Edwards Batch-only
  - #1 Siebel CRM
  - #1 8-chip SAP-SD 2-tier
  - #1 Oracle FLEXCUBE UBS
- **Two #1s Analytics**
  - #1 Oracle TimesTen, #1 Oracle OLAP
- **Eight #1s on SPEC CPU benchmarks**

SPARC T5
SPARC M5

Leads in every area!

ERP

HCM
CRM
SRM

SCM
FMS
BI-DW

OLTP

ORACLE

See benchmark disclosure slide

Objectives of the Next Oracle Processor

- Expand the high end of large mission-critical data servers
  - Large scale consolidation of virtualized applications
  - Large in-memory database and applications
  - Scaling to very high thread counts
  - High degree of robustness and application uptime
  - Software compatible – effortless workload migration

ORACLE

The Next Oracle Processor: SPARC M6

	nm	Cores	Threads	L3$	Memory per Socket	PCIe	Max. Sockets
T4	40	8	64	4MB	0.5TB	2*G2	4
T5	28	16	128	8MB	0.5TB	2*G3	8
M5	28	6	48	48MB	1TB	2*G3	32
M6	28	12	96	48MB	1TB	2*G3	96

ORACLE

Outline

- Lineage
- Features
- Scaling
- Reliability, Availability, Serviceability
- Summary

SPARC S3 Core

- Dual-issue, out-of-order

- Integrated encryption acceleration instructions

- Enhanced instruction set to accelerate Oracle SW stack
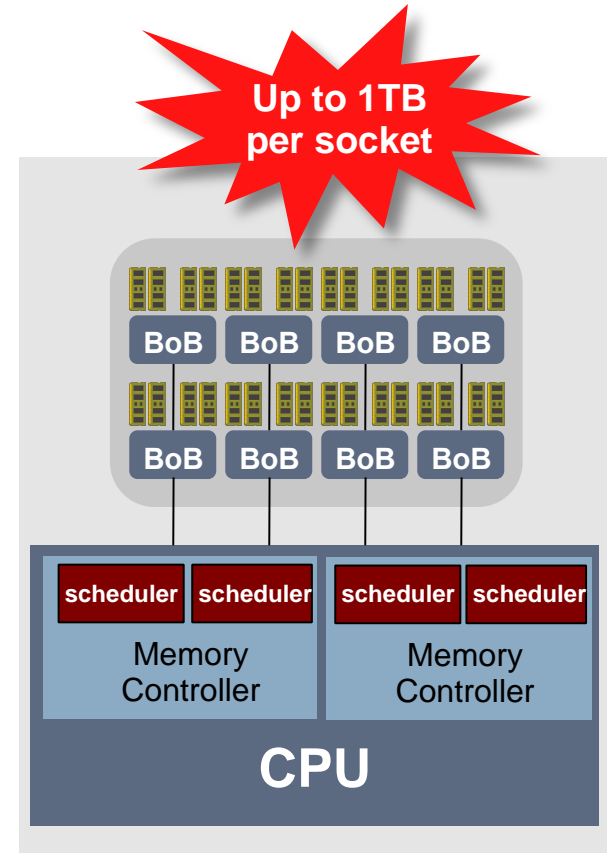
- 1-8 strands, dynamically threaded pipeline
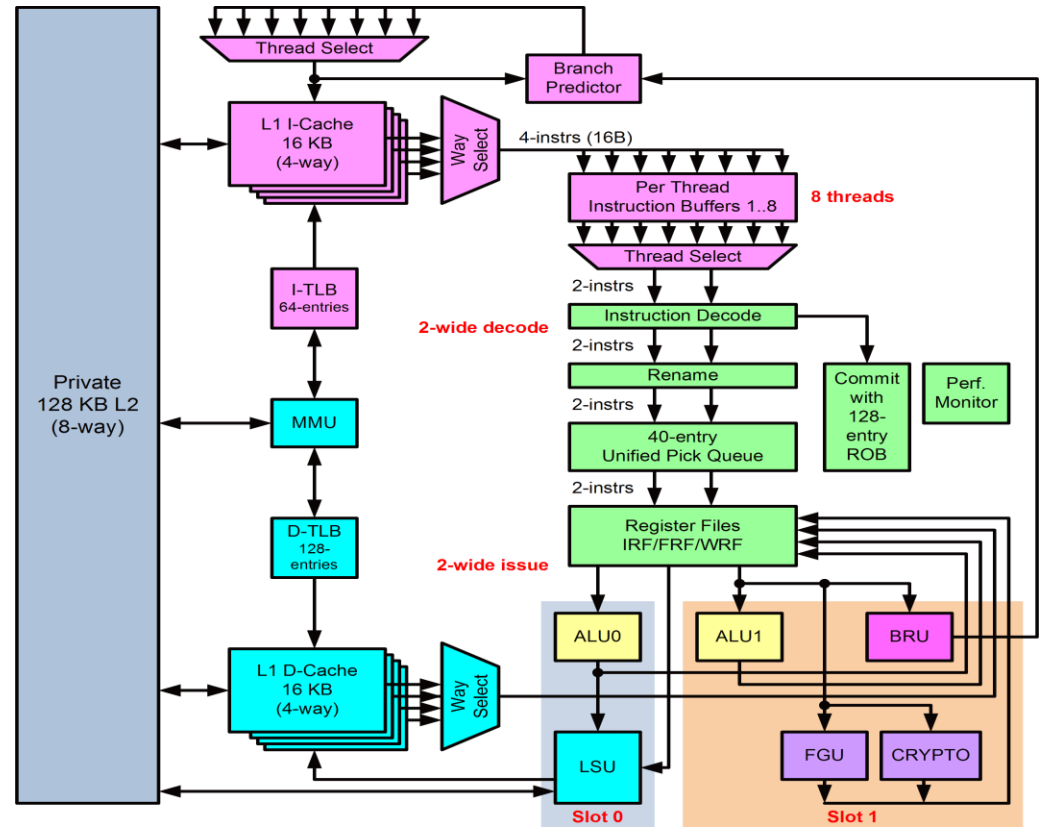
SPARC M6: Caches

- Per-core L1 and L2 Caches
  - L1 I-Cache: 16KB, 4-ways, 32-byte lines
  - L1 D-Cache: 16KB, 4-ways, 16-byte lines, write-through
  - L2 Unified Cache: 128KB, 8-ways, 32-byte lines, write-back, inclusive
- Per-chip Shared L3 Cache
  - 48MB, 4-banks,12-ways, 64-byte lines, inclusive, MOESI
  - Allocating DMA based on PCIe TLP Processing Hints
  - Request bundling to improve performance of high-access shared regions of database
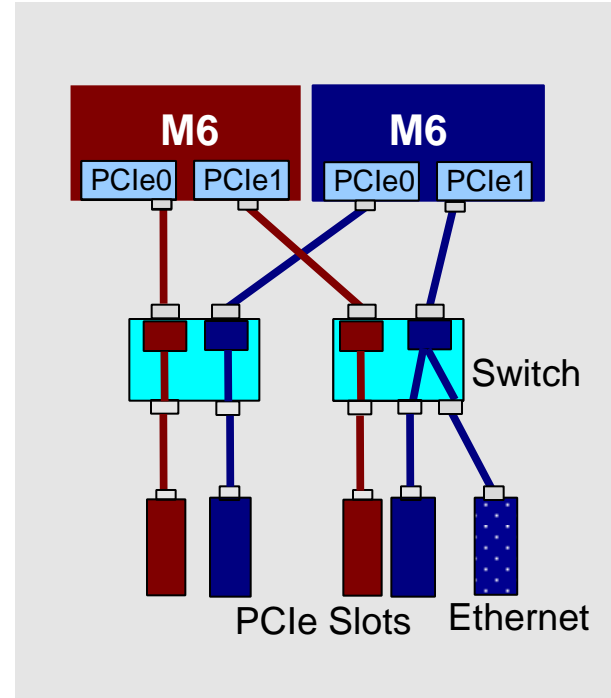
ORACLE

SPARC M6: Memory Subsystem

- Tuned for In-Memory Database and Applications

- Four high efficiency DDR3 schedulers
  - 16 DDR Channels per socket
  - Per rank scheduling
  - Dynamic adjustment of write vs. read priority
  - DIMM power saving modes

- Wide palette of address interleave settings to optimize the balance of performance, serviceability and power

Up to 1TB per socket

BoB BoB BoB BoB

BoB BoB BoB BoB

scheduler scheduler   scheduler scheduler

Memory Controller   Memory Controller

CPU

ORACLE

SPARC M6: IO Subsystem

- Dual x8 PCIe Gen3
- Atomic operations: fetch-add, swap, etc.
- TLP hints to direct DMA writes to L3
- PCIe power management
- Acceleration functions for virtual IO
- PCIe architected errors
- Error signaling via PCIe messages
- Support for independent reset of each PCIe root complex and the attached fabric

ORACLE

SPARC M6 Processor

Copyright © 2013, Oracle and/or its affiliates. All rights reserved.

ORACLE
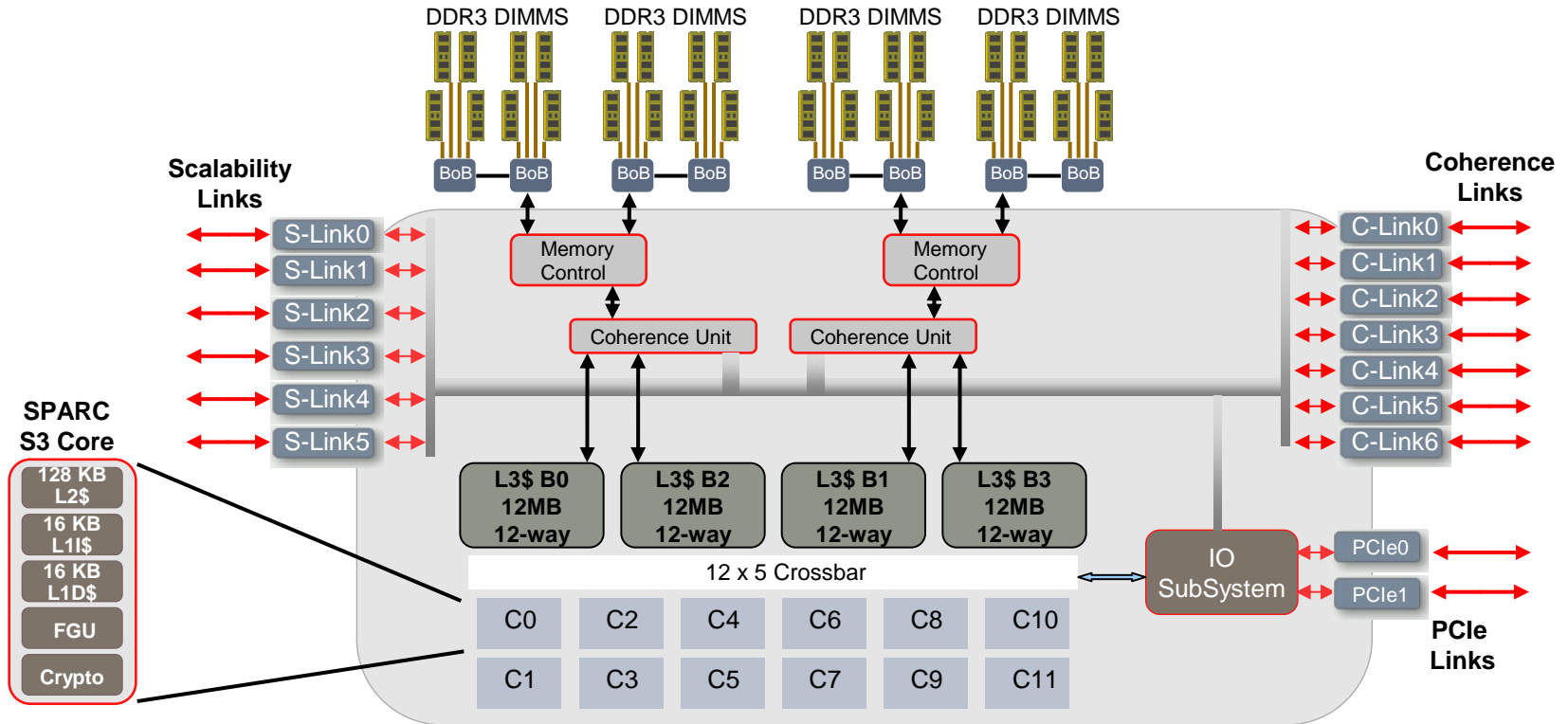
SPARC M6: Processor Overview

- 12 SPARC S3 cores, 96 threads
- 48MB shared L3 cache
- 4 DDR3 schedulers, maximum of 1TB of memory per socket
- 2 PCIe 3.0 x8 lanes
- Up to 8 sockets glue-less scaling
- Up to 96 sockets glued scaling
- 4.1 Tbps total link bandwidth
- 4.27 billion transistors
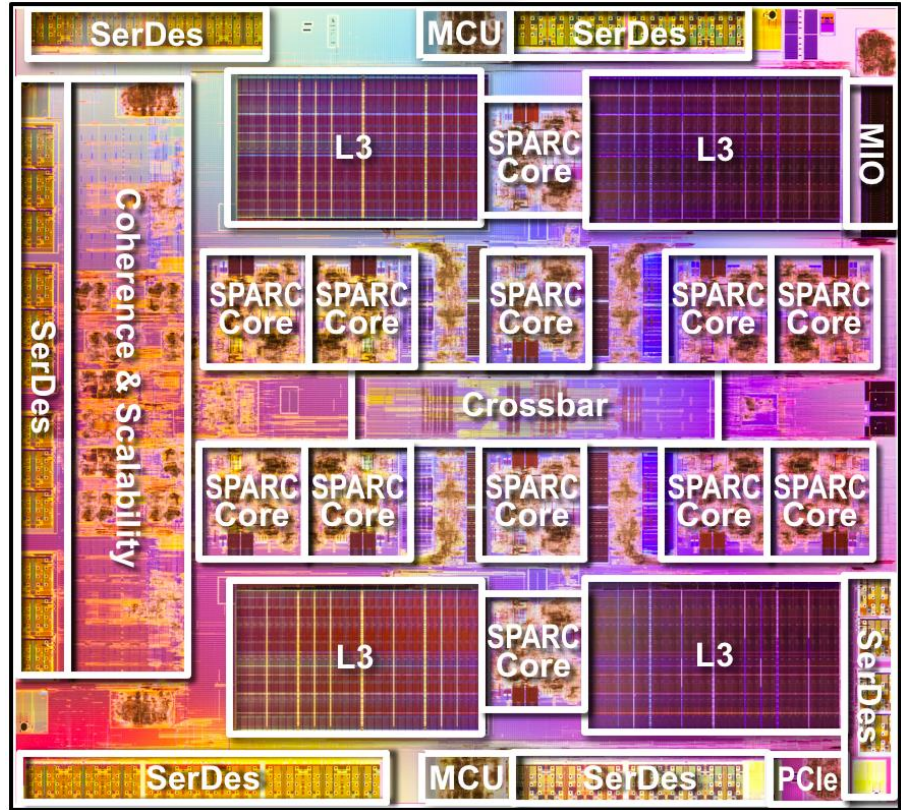
ORACLE

Outline

- Lineage
- Features
- Scaling
- Reliability, Availability, Serviceability
- Summary

ORACLE

SPARC M6 Scaling: Tightly Coupled SMP

- Up to eight processors directly connected using C-Links

- Can build 2-way and 4-way with multiple links between processors

- Can operate with de-configured sockets

- Memory and directory address sliced among the processors

- Different address hashing for memory home vs. directory home

ORACLE
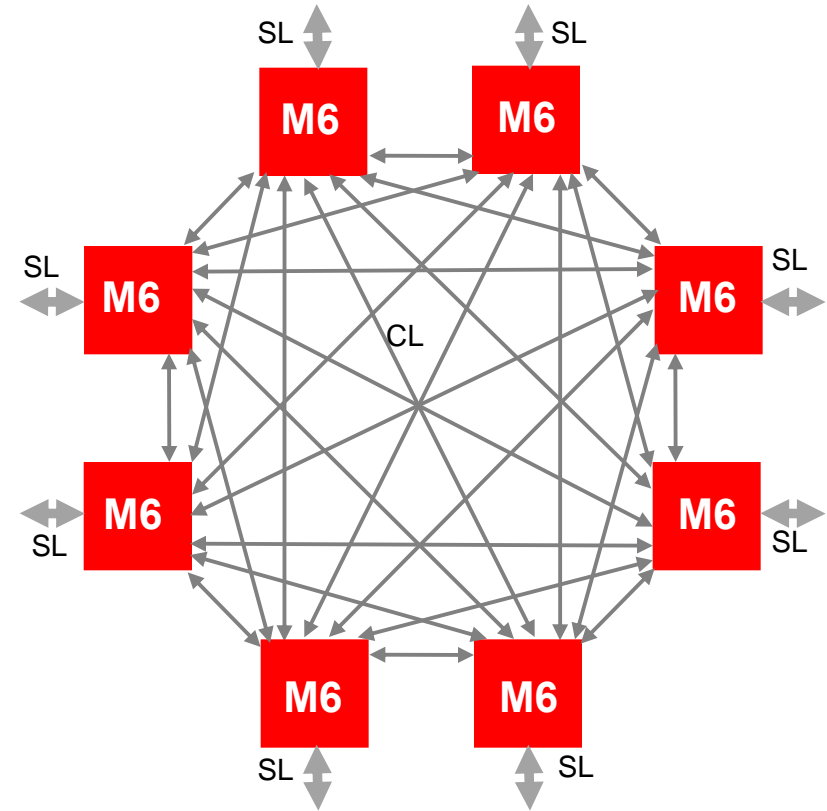
SPARC M6 Scaling: Beyond an SMP

- System directory is located in Bixby, connected to the processor via S-Links

- The 48 S-Links of an SMP represent different address planes

- Requests must first use C-Links to get to the proper processor, and then S-Links to get to the directory

- Responses back-track the path of the request

- Data dynamically routed across all available planes

ORACLE
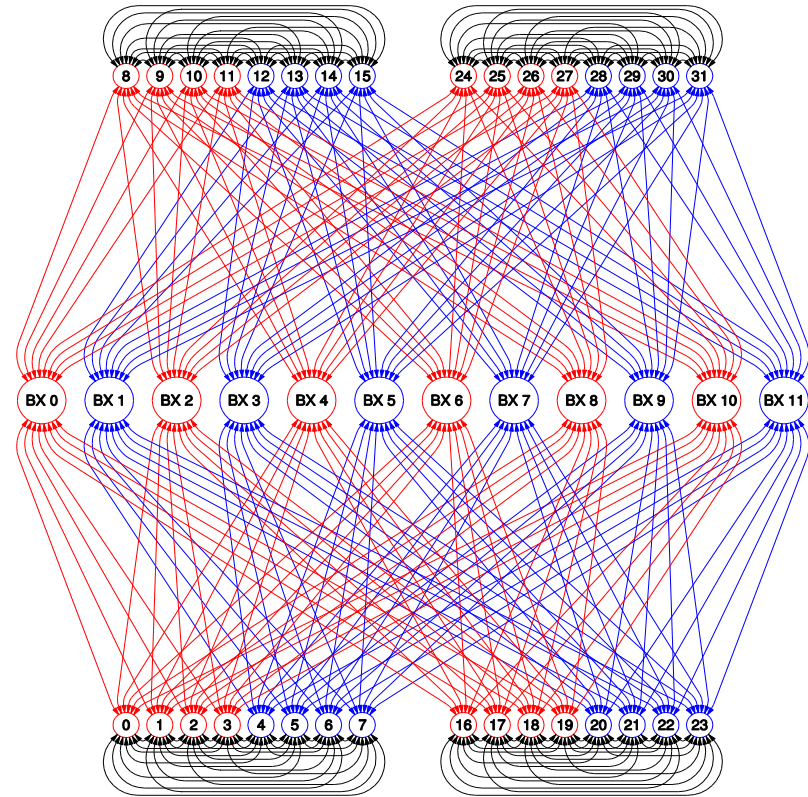
SPARC M6 Scaling: Larger Systems

ORACLE

ORACLE
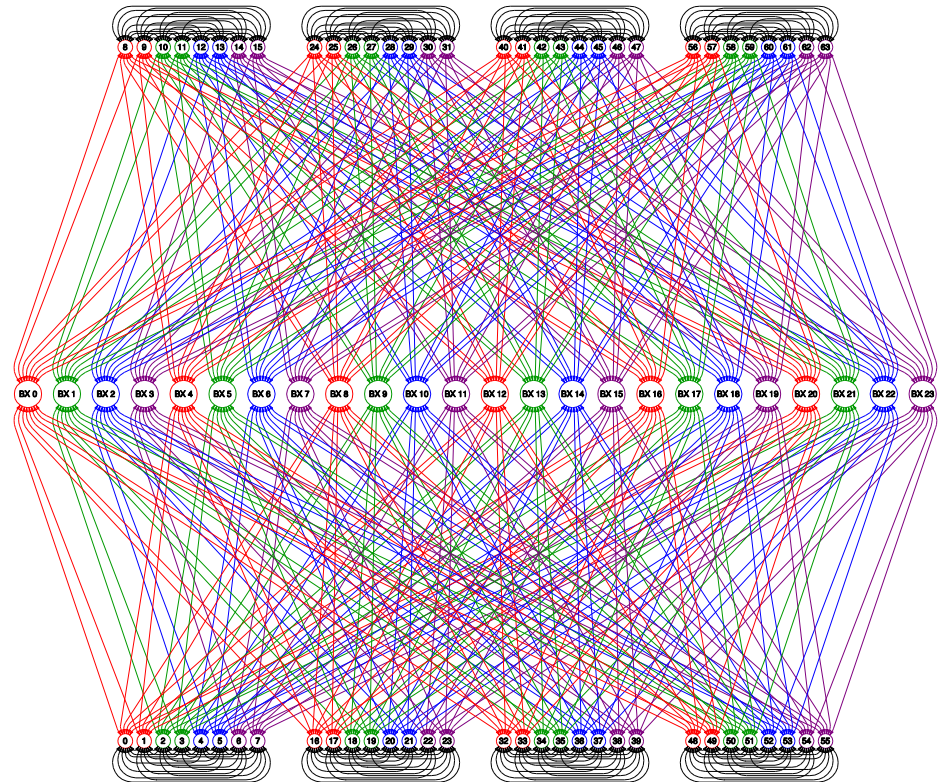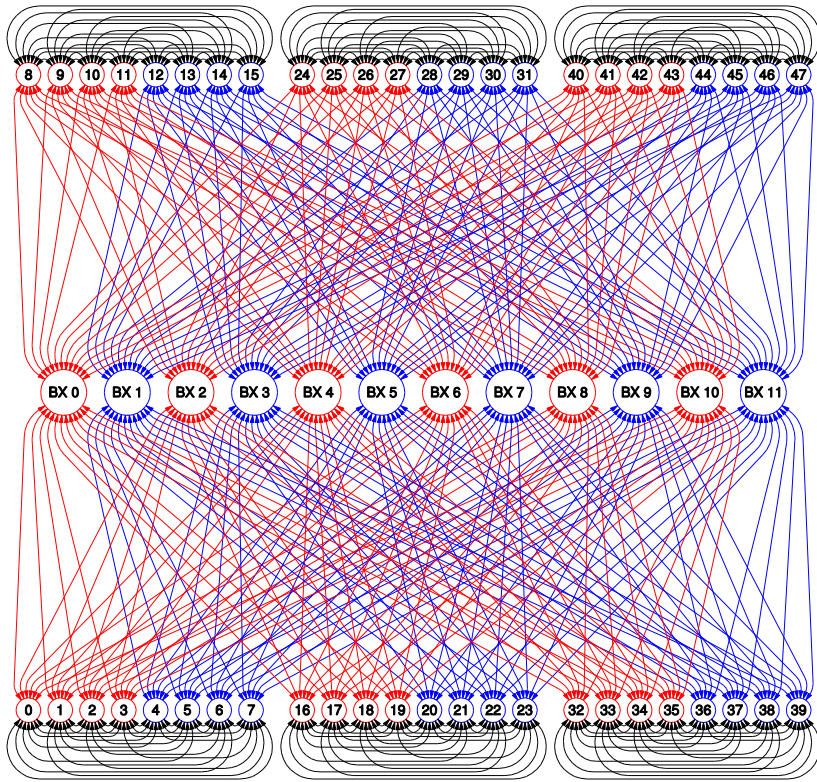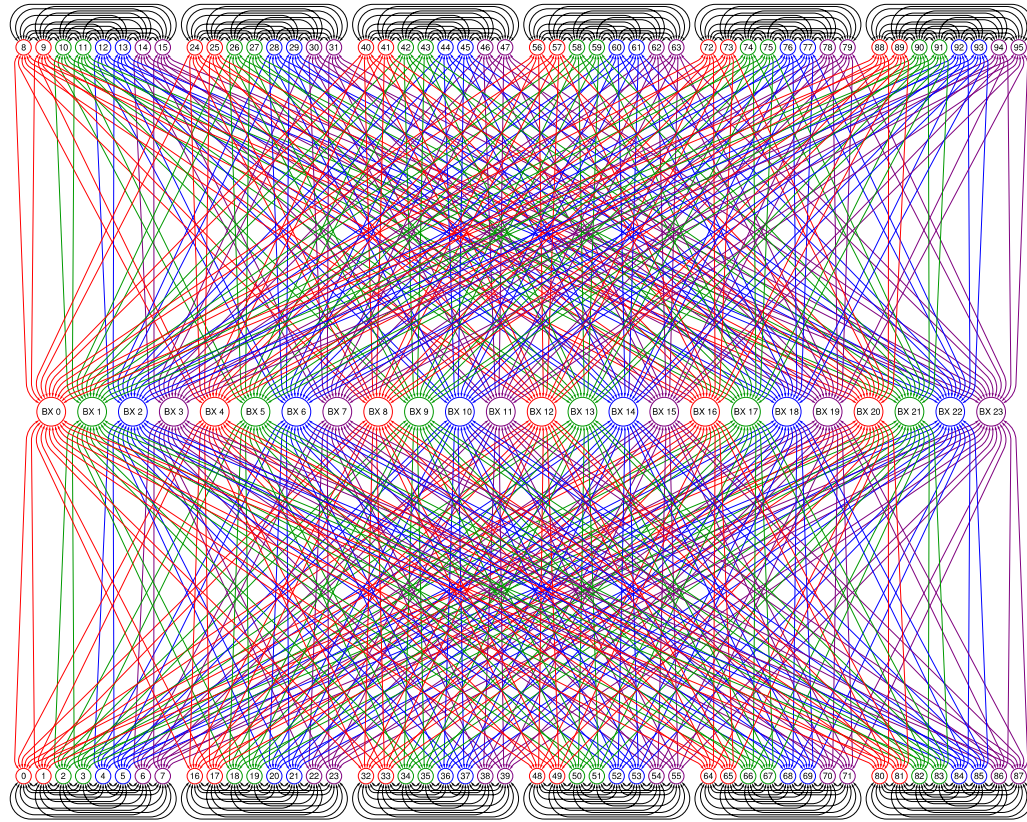
# SPARC M6 Scaling: Challenges of Coherence

- Small scale is already constrained
  - Trade-off between bandwidth, latency, and complexity
  - Intersection of ordering rules for SPARC and PCIe
- Large scale compounds the problem
  - Thousands of requesters (threads and IO)
  - Fully sized buffers not practical
  - Point-to-point connections not practical
  - Path asymmetries start to factor into the design

SPARC M6 Coherence: Simple Transaction

- Request is sent to the Directory SpecRead is sent to Memory Home

- Directory checks line state and returns ReqAck with what to expect

- If line is not in any cache, Directory tells Memory Home to source data

- If line is in cache(s), Directory tells a cache to source and if necessary, tells others to invalidate their copies

- Requester collects InvAck from all holders and then informs the Directory to unlock the line

| |

ORACLE
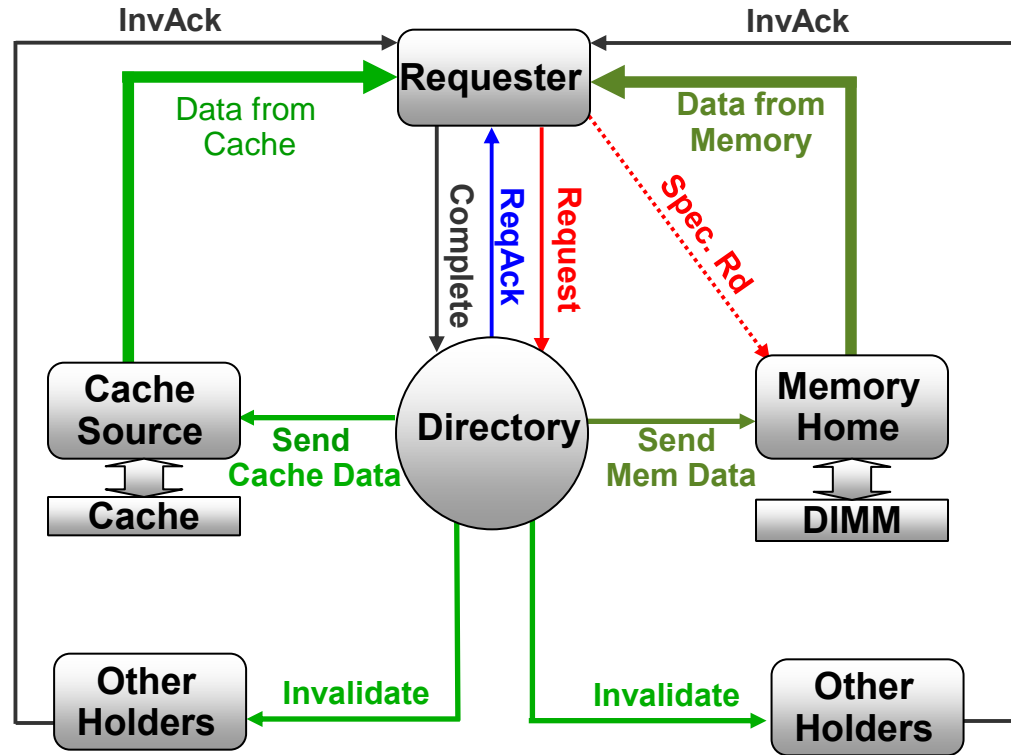
SPARC M6 Scaling: Implementation

- Measured pace of adoption of complexity
  - Small system coherence cannot scale up
  - M6 coherence architected for large-scale
  - Parameterized features allow efficient scale-down for smaller systems
  - Deployment from small to large
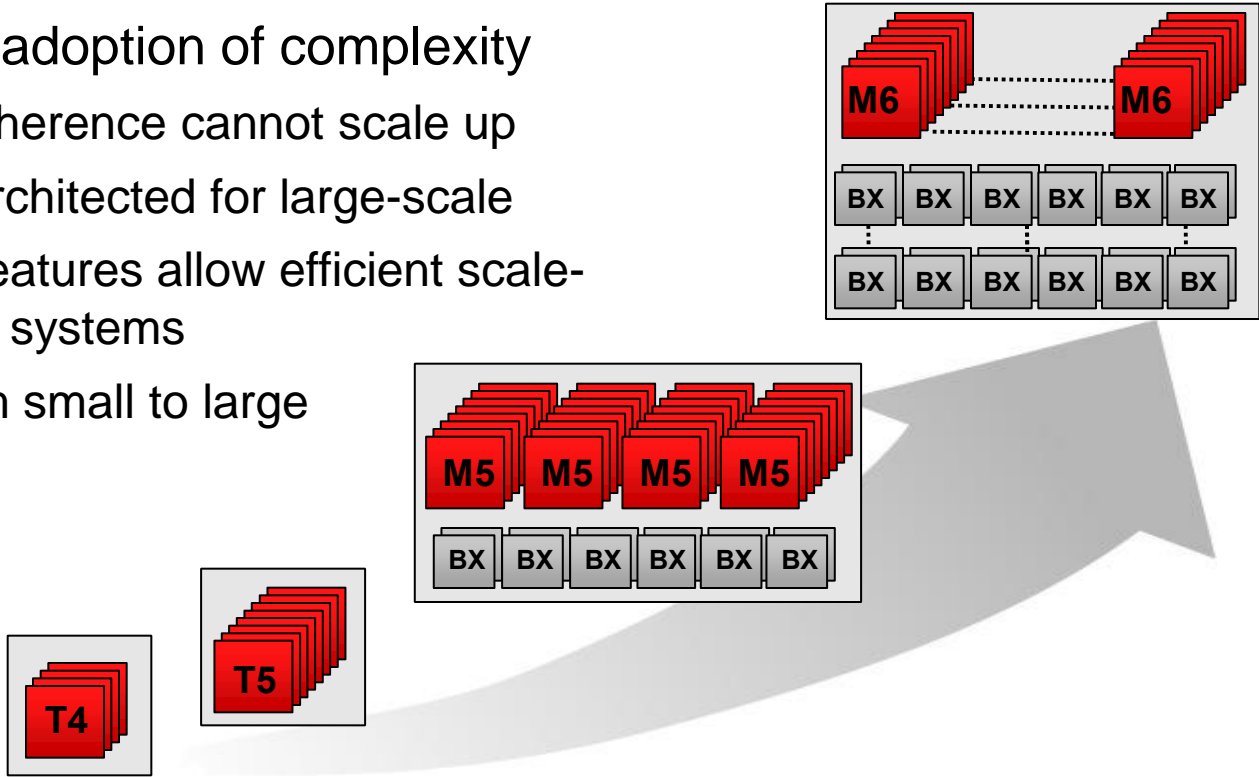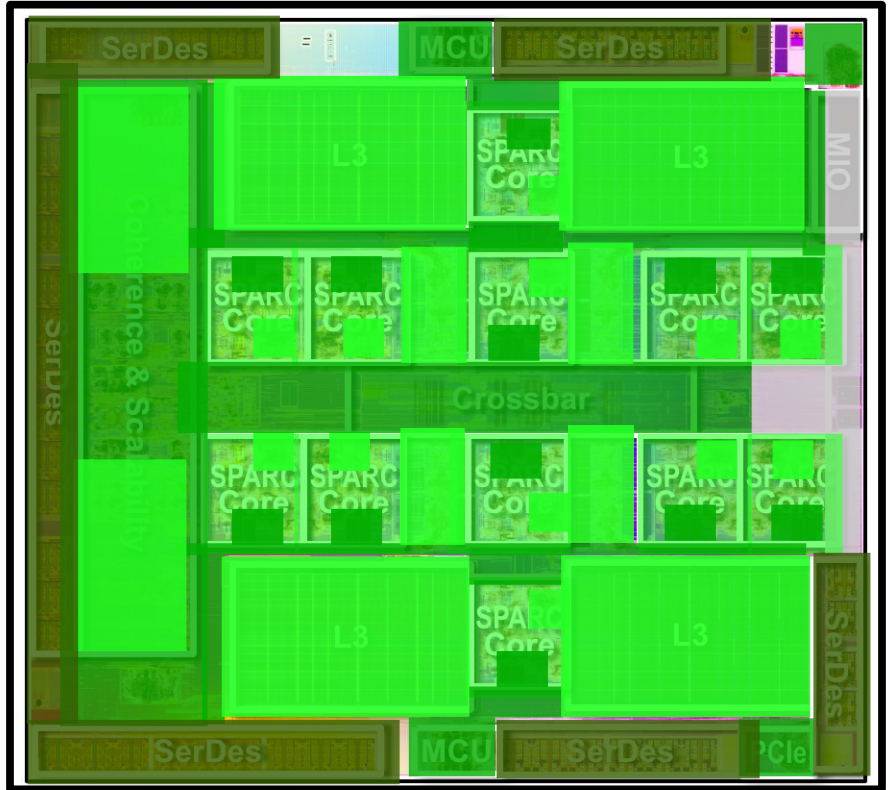
ORACLE

Outline

- Lineage
- Features
- Scaling
- Reliability, Availability, Serviceability
- Summary

SPARC M6 RAS: End-to-End Protection

- Internal Logic: parity and ECC

  – Architectural Registers

  – Cache structures

  – Internal networks

- Links: CRC retry

- CRC
- ECC with line retire
- Data-ECC, Address-parity
- Other (Parity, Retry etc)
- DFT, Debug etc.

ORACLE
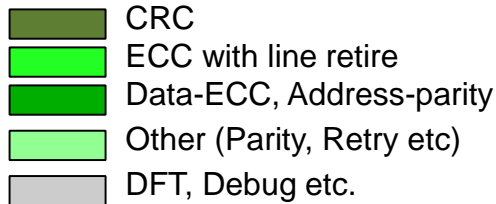
SPARC M6 RAS: Handling of Internal Errors

- In-line correction where possible, flush-and-retry if timing critical

- Auto-discard clean data, poison dirty data

  – Contain the error within a thread

- For cache structures

  – Retire the line aggressively and un-retire after detailed analysis

  – Use bypass path for replay to guarantee forward progress in presence of persistent errors

ORACLE

SPARC M6 RAS: Memory Errors

- ECC optimized for device failures

- Inline correction and auto write-back

- "Scrubber" prevents accumulation of upsets

- "E-retry" characterizes soft vs persistent

- Cell or word-line fail: Solaris retires page(s)

- Bit-line or pin fail: firmware deploys DIMM spare column

- Device fail: inline correction
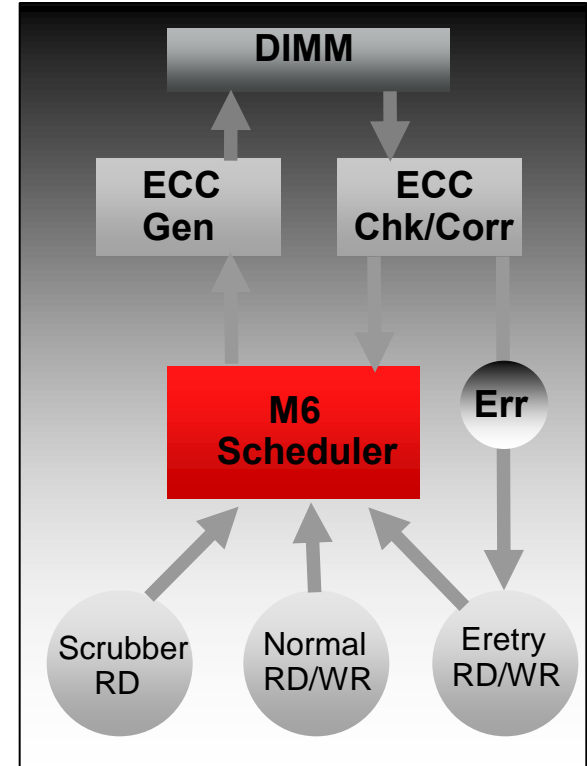
ORACLE

SPARC M6 RAS: General Error Handling Flow

Hardware

Detection and Clean-up
Gather Signature
De-configure Resources

Hypervisor

Assist with clean-up
Collect Hardware Data
Generate Report for SP
Manage De-configuration

Service Processor

Analyze Hypervisor Report
Update Error History
Initiate Service Call
Initiate De-configuration

Solaris

De-configure
User-Visible Resource

Offline SerDes Lanes

Retire Cache Lines

Retire Threads Retire Cores Retire Pages

Activate DIMM Spare Column

ORACLE

Outline

- Lineage
- Features
- Scaling
- Reliability, Availability, Serviceability
- Summary

ORACLE

SPARC M6: Summary

- At the leading edge of design and technology

- Tuned for Oracle workloads

- Extreme scaling and Best of Class RAS

- Enables Oracle's next Enterprise System

- Provides unprecedented level of performance for Oracle software stack and In-Memory Database and Applications

Applications
Middleware
Database
Operating System
Virtual Machine
Servers
Storage

ORACLE

Benchmark Disclosure Statement (1 of 2)

▪TPC Benchmark C, tpmC, and TPC-C are trademarks of the Transaction Processing Performance Council (TPC). SPARC T5-8 (8/128/1024) with Oracle Database 11g Release 2 Enterprise Edition with Partitioning, 8,552,523 tpmC, $0.55 USD/tpmC, available 9/25/2013, New Order 90th% Response Time 0.410sec. IBM Power 780 Cluster (24/192/768) with DB2 ESE 9.7, 10,366,254 tpmC, $1.38 USD/tpmC, available 10/13/2010, New Order 90th% Response Time 2.10 sec. IBM x3850 X5 (4/40/80) with DB2 ESE 9.7, 3,014,684 tpmC, $0.59 USD/tpmC, available 7/11/2011. IBM x3850 X5 (4/32/64) with DB2 ESE 9.7, 2,308,099 tpmC, $0.60 USD/tpmC, available 5/20/2011. IBM Flex x240 (2/16/32) with DB2 ESE 9.7, 1,503,544 tpmC, $0.53 USD/tpmC, available 8/16/2012. IBM Power 780 (2/8/32) with IBM DB2 9.5, 1,200,011 tpmC, $0.69 USD/tpmC, available 10/13/2010. Source: http://www.tpc.org/tpcc, results as of 3/26/2013.

▪ SPEC and the benchmark name SPECjEnterprise are registered trademarks of the Standard Performance Evaluation Corporation. Results from www.spec.org as of 3/26/2013. SPARC T5-8, 57,422.17 SPECjEnterprise2010 EjOPS; SPARC T4-4, 40,104.86 SPECjEnterprise2010 EjOPS; Sun Server X2-8, 27,150.05 SPECjEnterprise2010 EjOPS; Cisco UCS B440 M2, 26,118.67 SPECjEnterprise2010 EjOPS; IBM Power 780, 16,646.34 SPECjEnterprise2010 EjOPS. IBM PowerLinux 7R2, 13,161.07 SPECjEnterprise2010 EjOPS. SPARC T3-4 9456.28 SPECjEnterprise2010 EjOPS. SPARC T5-8 (SPARC T5-8 Server base package, 8xSPARC T5 16-core processors, 128x16GB-1066 DIMMS, 2x600GB 10K RPM 2.5. SAS-2 HDD, 4x Power Cables) List Price $268,742. IBM Power 780 (IBM Power 780:9179 Model MHB, 8x3.86GHz 16-core, 64x one processor activation, 4xCEC Enclosure with IBM Bezel, I/O Backplane and System Midplane,16x 0/32GB DDR3 Memory (4x8GB) DIMMS-1066MHz Power7 CoD Memory, 12x Activation of 1 GB DDR3 Power7 Memory, 5x Activation of 100GB DDR3 Power7 Memory, 1x Disk/Media Backplane. 2x 146.8GB SAS 15K RPM 2.5. HDD (AIX/Linux only), 4x AC Power Supply 1725W) List Price $992,023. Source: Oracle.com and IBM.com, collected 03/18/2013. SPEC and the benchmark name SPECjEnterprise are registered trademarks of the Standard Performance Evaluation Corporation. Results from www.spec.org as of 5/1/2013. SPARC T5-8, 27,843.57 SPECjEnterprise2010 EjOPS; IBM Power 780, 10,902.30 SPECjEnterprise2010 EjOPS. Oracle server only hardware list price is $298,494 and total hardware plus software list price is $1,565,092 http://www.oracle.com as of 4/24/2013. IBM server only HW list price is $835,555 and HW+SW cost of $2,174,152.00 and BM PowerLinux 7R2 server total hardware plus software cost of $819,451.00 based on public pricing from http://www.ibm.com as of 4/24/2013.

▪SPEC & the benchmark name SPECjbb are registered trademarks of Standard Performance Evaluation Corporation (SPEC). Results as of 3/26/2013, see http://www.spec.org for more information. SPARC T5-2 75,658 SPECjbb2013-MultiJVM max-jOPS, 23,334 SPECjbb2013-MultiJVM critical-jOPS. Sun Server X2-4 65,211 SPECjbb2013-MultiJVM max-jOPS, 22,057 SPECjbb2013-MultiJVM critical-jOPS. Sun Server X3-2 41,954 SPECjbb2013-MultiJVM max-jOPS, 13,305 SPECjbb2013-MultiJVM critical-jOPS. SPARC T4-2 34,804 SPECjbb2013-MultiJVM max-jOPS, 10,101 SPECjbb2013-MultiJVM critical-jOPS. HP ProLiant DL560p Gen8 66,007 SPECjbb2013-MultiJVM max-jOPS, 16,577 SPECjbb2013-MultiJVM critical-jOPS. HP ProLiant ML350p Gen8 40,047 SPECjbb2013-MultiJVM max-jOPS, 12,308 SPECjbb2013-MultiJVM critical-jOPS. Supermicro X8DTN+ 20,977 SPECjbb2013-MultiJVM max-jOPS, 6,188 SPECjbb2013-MultiJVM critical-jOPS. HP ProLiant ML310e Gen8 12,315 SPECjbb2013-MultiJVM max-jOPS, 2,908 SPECjbb2013-MultiJVM critical-jOPS. Intel R1304BT 6,198 SPECjbb2013-MultiJVM max-jOPS, 1,722 SPECjbb2013-MultiJVM critical-jOPS, HP DL980 G7 106,141 SPECjbb2013-MultiJVM max-jOPS, 23268 SPECjbb2013-MultiJVM critical-jOPS; Fujitsu SPARC M10-4S 4-chip 3GHz SPARC64 X, SPECjbb2013-Multi-JVM 83,909 max-jOPS, 50,562 SPECjbb2013-Multi-JVM critical-jOPS.

ORACLE

Benchmark Disclosure Statement (2 of 2)

- Copyright 2013, Oracle &/or its affiliates. All rights reserved. Oracle & Java are registered trademarks of Oracle &/or its affiliates.Other names may be trademarks of their respective owners..
- Two-tier SAP Sales and Distribution (SD) Standard Application benchmarks SAP Enhancement package 5 for SAP ERP 6.0 as of 3/26/13:SPARC M5-32 (32 processors, 192 cores, 1536 threads) 85,050 SAP SD users, 32 x 3.6 GHz SPARC M5, 4 TB memory, Oracle Database 11*g*, Oracle Solaris 11, Cert# 2013009.   SPARC T5-8 (8 processors, 128 cores, 1024 threads) 40,000 SAP SD users, 8 x 3.6 GHz SPARC T5, 2 TB memory, Oracle Database 11g, Oracle Solaris 11, Cert# 2013008. IBM Power 760 (8 processors, 48 cores, 192 threads) 25,488 SAP SD users, 8 x 3.41 GHz IBM POWER7+, 1024 GB memory, DB2 10, AIX 7.1, Cert#2013004. Two-tier SAP Sales and Distribution (SD) Standard Application benchmarks SAP Enhancement package 4 for SAP ERP 6.0 as of 4/30/12:IBM Power 795 (32 processors, 256 cores, 1024 threads) 126,063 SAP SD users, 32 x 4 GHz IBM POWER7, 4 TB memory, DB2 9.7, AIX7.1, Cert#2010046. SPARC Enterprise Server M9000 (64 processors, 256 cores, 512 threads) 32,000 SAP SD users, 64 x 2.88 GHz SPARC64 VII, 1152 GB memory, Oracle Database 10g, Oracle Solaris 10, Cert# 2009046. SAP, R/3, reg TM of SAP AG in Germany and other countries. More info www.sap.com/benchmark
- SPEC & benchmark names SPECfp, SPECint are registered trademarks of the Standard Performance Evaluation Corporation. Results as of March 26, 2013 from www.spec.org and this report. SPARC T5-8: 3750 SPECint_rate2006, 3490 SPECint_rate_base2006, 3020 SPECfp_rate2006, 2770 SPECfp_rate_base2006; SPARC T5-1B: 467 SPECint_rate2006, 436 SPECint_rate_base2006, 369 SPECfp_rate2006, 350 SPECfp_rate_base2006. IBM Power 780 8-chip 3.92GHz: 2640 SPECint_rate2006. IBM Power 710 Express 1-chip 3.556GHz: 289 SPECint_rate2006.
- TPC Benchmark, TPC-H, QphH, QthH, QppH are trademarks of the Transaction Processing Performance Council (TPC). Results as of 6/7/13, prices are in USD. SPARC T5-4 www.tpc.org/3288; SPARC T4-4 www.tpc.org/3278; SPARC Enterprise M9000 www.tpc.org/3262; SPARC Enterprise M9000 www.tpc.org/3258; IBM Power 780 www.tpc.org/3277; HP ProLiant DL980 www.tpc.org/3285.

ORACLE

Glossary

- MOESI – Modified-Owned-Exclusive-Shared-Invalid
- SEC-DED – Single-bit Error Correcting - Double-bit Error Detecting
- BoB – Buffer on Board
- CRC – Cyclic Redundancy Check
- ECC – Error Correcting Code
- SMP – Shared Memory Processor
- RAS – Reliability Availability Servicability
- TLP – Transaction Layer Packet (PCIe)

Hardware and Software

ORACLE®

Engineered to Work Together