Oracle OpenWorld 2019

Safe Harbor

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, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

Statements in this presentation relating to Oracle's future plans, expectations, beliefs, intentions and prospects are "forward-looking statements" and are subject to material risks and uncertainties. A detailed discussion of these factors and other risks that affect our business is contained in Oracle's Securities and Exchange Commission (SEC) filings, including our most recent reports on Form 10-K and Form 10-Q under the heading "Risk Factors." These filings are available on the SEC's website or on Oracle's website at http://www.oracle.com/investor. All information in this presentation is current as of September 2019 and Oracle undertakes no duty to update any statement in light of new information or future events.

ORACLE

Introducing Exadata X8M

In-Memory Performance with All the
Benefits of Shared Storage for both OLTP
and Analytics

Bob Thome

VP, Product Management

New RDMA Fabric
Persistent Memory

Exadata Vision

Dramatically Better Platform for All Database Workloads

- **Ideal Database Hardware** – scale-out, database optimized compute, networking, and storage for fastest performance and lowest cost

- **Smart System Software** – specialized algorithms vastly improve all aspects of database processing: OLTP, Analytics, Consolidation

- **Automated Management** – automation and optimization of configuration, updates, performance, and management culminating in Fully Autonomous Infrastructure and Database

Thousands of Critical Deployments Since 2008

Financial Services, Telecoms, Healthcare, Retail, Public Sector, Travel, Manufacturing, Professional Services, Consumer Goods, Education, Utilities, …
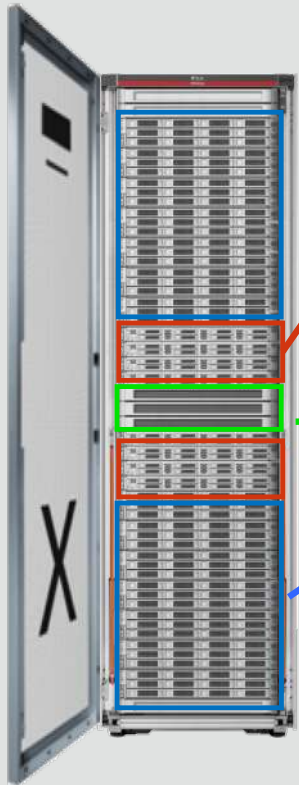
Best for ALL workloads

- Petabyte Warehouses
- Super Critical Systems
  - Financial trading
  - Process manufacturing
  - E-commerce
- Packaged Applications
  - SAP, Oracle, Siebel, PSFT, …
- Database Consolidation

77% of Fortune Global 100 Run Exadata

Exadata X8 Hardware (changes from X7 in red)

- Shipping since March 2019

- Scale-Out 2-Socket or 8-Socket Database Servers
  - 15% faster clock – latest 24 core Intel Cascade Lake
  - Spectre & Meltdown mitigated in silicon

- 40 Gb/sec Unified InfiniBand Internal Fabric

- Scale-Out *intelligent* 2-Socket Storage Servers
  - 60% more cores to offload database processing
    - Latest 16 core Intel Cascade Lake CPUs
  - 40% higher capacity 14 TB disk drives

Database Server

High-Capacity (HC) Storage

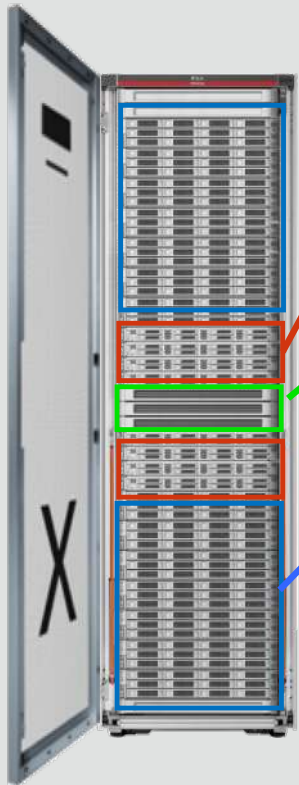**Extreme Flash (EF) Storage**

Extended (XT) Storage

Exadata X8M

Continues Tradition of
State-of-the-Art Hardware

Exadata X8M (changes from X8 in red)

- Scale-Out 2 or 8 Socket Database Servers
  - Latest **24 core** Intel Cascade Lake CPUs

- 100 Gb/sec RDMA over Converged Ethernet RoCE internal fabric

- Scale-Out *intelligent* 2-Socket Storage Servers
  - 1.5 TB Persistent Memory per storage server
  - Three tiers of storage: PMEM, NVMe, HDD

- Enhanced consolidation using Linux KVM

Database Server

High-Capacity (HC) Storage

Extreme Flash (EF) Storage

Extended (XT) Storage

In-Memory Analytics Performance in Shared Storage

- Exadata Flash throughput approaches DRAM throughput
  - Analytics SQL bottleneck moves from I/O to CPU

- Exadata storage automatically transforms table data into In-Memory DB columnar formats in Exadata Flash cache
  - Enables fast CPU vector processing in storage server queries

- Uniquely optimizes next generation flash as memory
  - Works for both row format OLTP databases, and for Hybrid Columnar Compressed Analytics databases

- New in Exadata Smart Software 19.3.0
  - Decryption only the Columns needed for a query
  - In-Memory Aggregation algorithms offloaded to Exadata

In-Memory Columnar Scans

In-Flash Columnar Scans
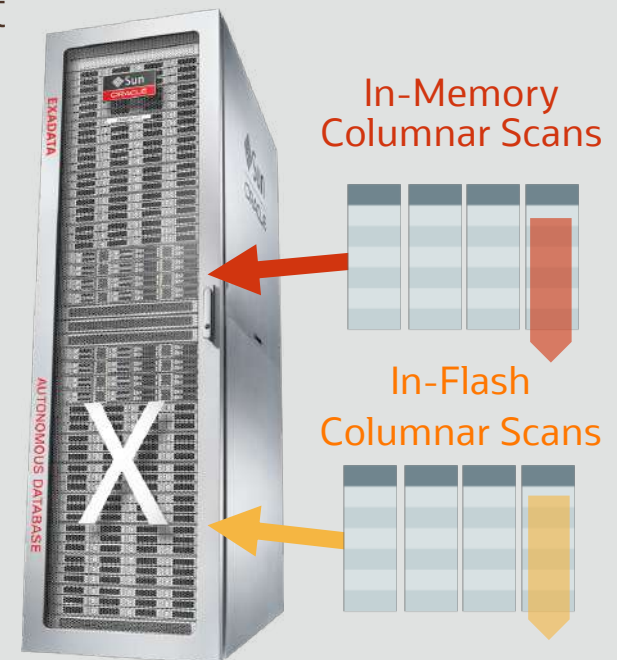
Exadata X8M Storage Performance

- Exadata X8M storage performs comparable to in-memory
  - With capacity, sharing, and cost benefits of shared storage
- 16 Million OLTP Read IOPS (8K I/Os)
  - 2.5x faster than Exadata X8
- < 19 microsecond OLTP I/O latency
  - 10x faster than Exadata X8
- Ultra fast log file writes to accelerate transactions
- 560 GB/sec Analytic Scan throughput
  - Over 1 TB/sec analytic scans with columnar data in flash
- Performance scales as more racks are added

Each rack has up to
3.0 PB Raw Disk
920 TB NVMe Flash
27 TB PMEM
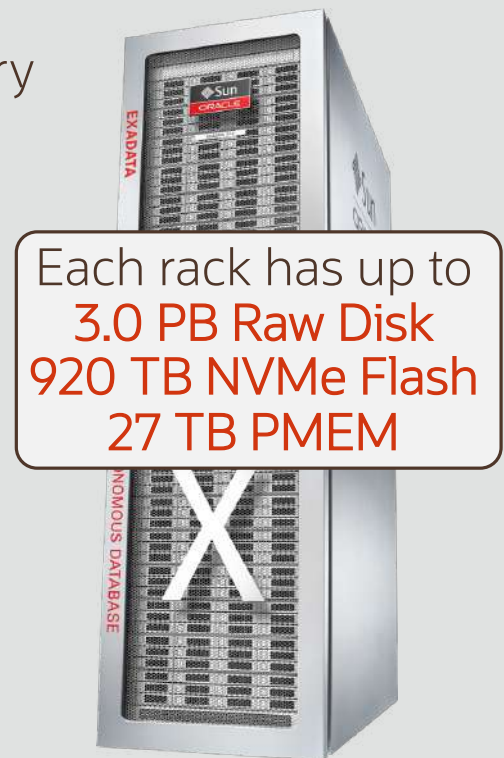
Exadata Smart System Software

- Fastest Analytics
  - Unique Smart Scan automatically offloads data intensive SQL operations to storage
  - Unique Smart Flash Cache and Storage Index automatically accelerate database I/O
  - Unique automatic conversion of data to fast In-Memory Columnar format in flash

- Fastest OLTP
  - Fastest OLTP I/O with scale-out storage, RDMA, and NVMe flash
  - Fastest scale-out with unique RDMA algorithms for inter-node cluster coordination
  - Fully redundant and fastest recovery from failed or sick components

- Best Consolidation
  - Uniquely prioritizes latency sensitive or important workloads through full stack
  - Uniquely isolates workloads from multiple tenants through full stack

Exadata X8M
# RoCE Networking

Exadata Uses RDMA for Extreme Performance

- With Remote Direct Memory Access (RDMA) one computer can access data from a remote computer without any OS or CPU involvement
  - Network card directly reads/writes memory with no extra copying or buffering and very low latency

- RDMA is integral to the Exadata high-performance architecture, enabling:
  - High throughput and low-CPU usage for large data transfers
  - Unique Direct-to-Wire Protocol to deliver 3x faster inter-node OLTP cluster messaging
  - Unique Smart Fusion Block Transfer that eliminates log write on inter-node block move
  - Unique RDMA protocol to coordinate transactions between nodes

RoCE – RDMA Over Converged Ethernet

- RDMA over Converged Ethernet (RoCE) is a protocol that runs InfiniBand RDMA software on top of Ethernet
  - Same software at upper levels of network protocol stack
  - InfiniBand packets sent as Ethernet UDP packets at low level

- RoCE on Exadata supports all Exadata RDMA optimizations

- RoCE enables scalability and volume of Ethernet with speed of RDMA

- Unifies Cluster switches and Client switches, saving costs

RoCE is Industry Standard
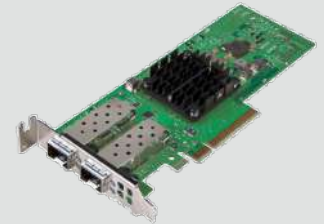
- Defined by an Open Consortium
  - InfiniBand Trade Association (IBTA)
  - Developed in open source and maintained in upstream Linux

- Supported by major network card vendors: Broadcom, Intel, Mellanox

- Supported by major switch vendors: Arista, Cisco, Juniper, Mellanox

- Exadata X8M uses Mellanox Card and Cisco switch

- RoCE also used by new storage products implementing remote access to flash drives on the network - NVMe (flash I/O) over Fabrics

New RoCE Internal Network Fabric

- InfiniBand was the only viable RDMA capable network at the inception of Exadata, but now Ethernet has caught up

- RoCE stands for RDMA over Converged Ethernet (RoCE)

- Industry-standard RDMA capable network cards, switches, protocol

- Exadata RoCE provides RDMA speed and reliability on Ethernet fabric
  - 100 Gb/sec throughput
  - Zero packet loss messaging
  - Prioritization of critical database messages
  - Latest KVM based virtualization

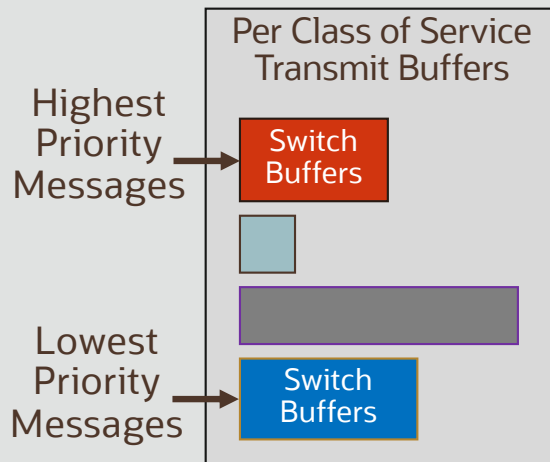**World's First and Only RoCE-based Database Machine**

RoCE Details

InfiniBand vs RoCE Ethernet

- InfiniBand was the only viable technology to implement RDMA at the inception of Exadata but now Ethernet has caught up

- RoCE matches previous InfiniBand advantages
  - Throughput - 100 Gb/sec Ethernet matches throughput of recent InfiniBand
  - Latency - Ethernet latency is now also very fast
  - RDMA protocols - protocols have now been ported on top of RoCE Ethernet
  - Lossless messaging - new RoCE ethernet switches and NICs achieve this

- InfiniBand disadvantages
  - Not common in enterprises, niche technology, not as virtualization friendly

# Exadata X8M RoCE – High Priority Networking

Network Switch

Per Class of Service
Transmit Buffers

Highest
Priority
Messages → Switch Buffers

Lowest
Priority
Messages → Switch Buffers

- Network prioritization for latency-sensitive DB algorithms
  - Ensures that messages requiring low latency are not slowed by high throughput messages
    - Low latency: cluster heartbeat, transaction commit, cache fusion
    - High throughput: backup, reporting, batch
- RoCE Class of Service (CoS)
  - Allows packets to be sent on multiple classes of service, each with separate network buffers for independence
- Exadata uniquely chooses the best Class of Service for each database message

Exadata X8M RoCE – Avoiding Packet Loss

Network Switch

Per Class of Service Transmit Buffers

Switch Buffers
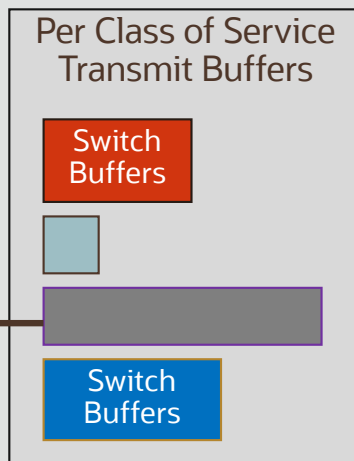
Switch Buffers

Buffer is Full so Pause
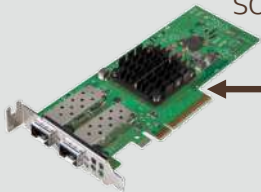
- Packet loss is usually caused by congestion
  - Packets sent faster than receiver or switches can process
  - Less common sources – switch failure, link failure
- Conventional Ethernet silently drops packets and expects retransmits if sends are too fast
  - Packet drop causes huge hit to latency and throughput
- Exadata RoCE avoids packet drops using:
- RoCE Priority-based Flow Control (PFC)
  - RoCE switch tells sender to pause if switch buffer is full
- RoCE Explicit Congestion Notification (ECN)
  - RoCE switch marks packet flow as too fast, telling source to slow down packet sends

Exadata X8M RoCE Instant Failure Detection
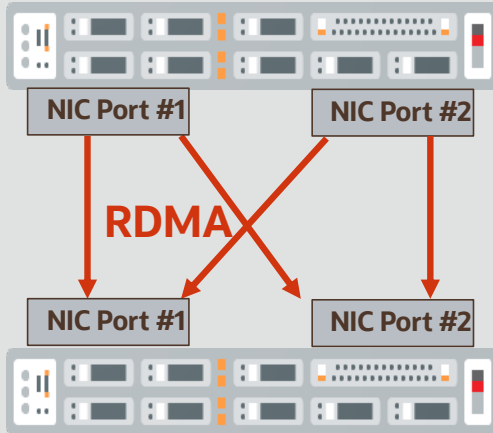
- Exadata uses frequent heartbeat messages between nodes to detect possible server failure

- Server failure detection normally requires long timeout to avoid false server evictions from cluster

  - Hard to quickly distinguish between whether a slow response to heartbeat is due to very high CPU load or to server failure

- Exadata uses RDMA to quickly confirm server failure

  - RDMA uses hardware, so remote ports respond even if software is slow

  - 4 RDMA Reads are sent to suspect server,

  - Across all combinations of source and target ports

  - If all 4 RDMAs fail, server is evicted from cluster

Exadata RoCE Delivers Extreme Performance

- Exadata RoCE internal fabric provides an extremely fast and low-latency connection for database and storage servers
- RoCE provides all the benefits previously unique to Infiniband, but with broader industry support for better integration in the data center
- Lower switch costs with unified private and client networks
- Exadata transparently prioritizes traffic by type, ensuring best performance for critical messages
- Exadata automatically optimizes network communications by ensuring packets are delivered on the first try without costly retransmissions
- Exadata eliminates stalls due to failures by immediately detecting server failures without waiting for lengthy timeouts

Exadata X8M

In-Memory Performance for OLTP
with all the benefits of Shared Storage

New Persistent Memory

- Persistent memory is a new silicon technology
  - Capacity, performance, and price are between DRAM and flash
- Intel® Optane™ DC Persistent Memory:
  - Reads at memory speed – much faster than flash
  - Writes survive power failure unlike DRAM
- Exadata implements sophisticated algorithms to maintain integrity of data on PMEM during failures
  - Call special instructions to flush data from CPU cache to PMEM
 - Complete or backout sequence of writes interrupted by a crash

Faster

Higher Cost Per GB

DRAM

Persistent Mem

FLASH
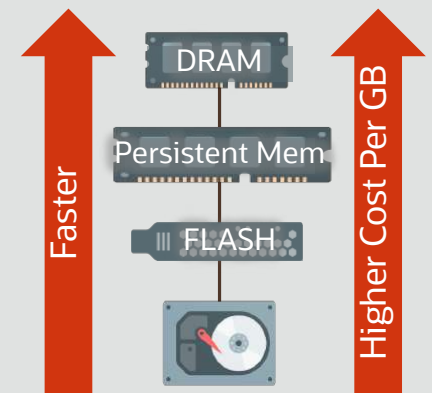
Persistent Memory with Traditional Storage

Compute Server

SAN

Storage Server

Hot
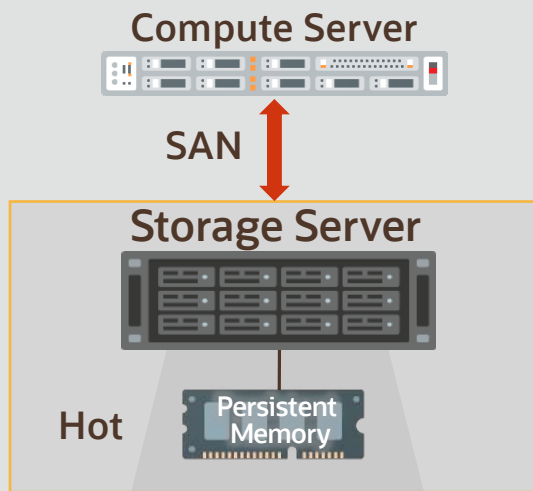
Persistent Memory

- Persistent Memory usage with traditional storage:
  - Database issues read I/O call to OS
  - OS sends message to storage
  - Storage CPU issues read to Persistent Memory
  - Storage CPU send reply to Server OS
  - Server OS wakes up Database
- Speed of Persistent Memory read is overwhelmed by high cost of network and I/O software, interrupts, and context switches
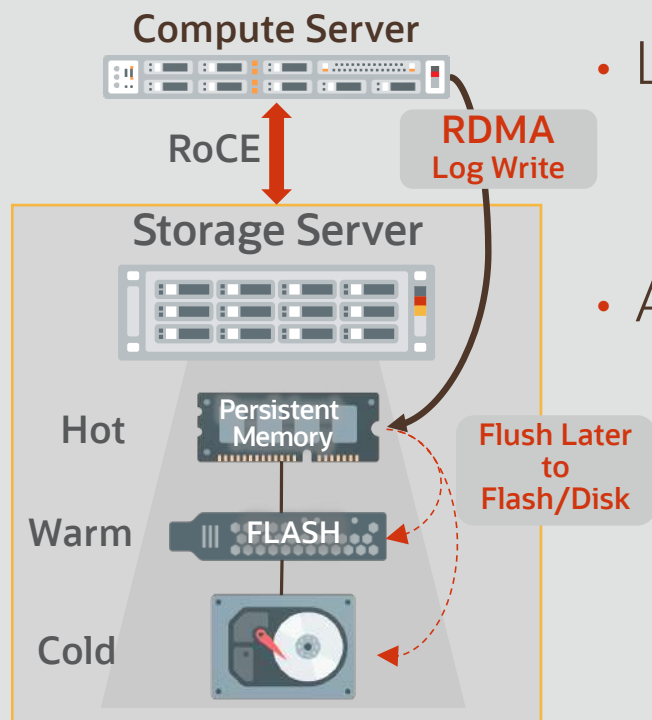- Very little benefit to users

Exadata X8M With Persistent Memory Data Accelerator

World's First and Only Shared Persistent Memory Optimized for Database

Compute Server

RoCE   RDMA

Storage Server

Hot   Persistent Memory

Warm   FLASH

Cold
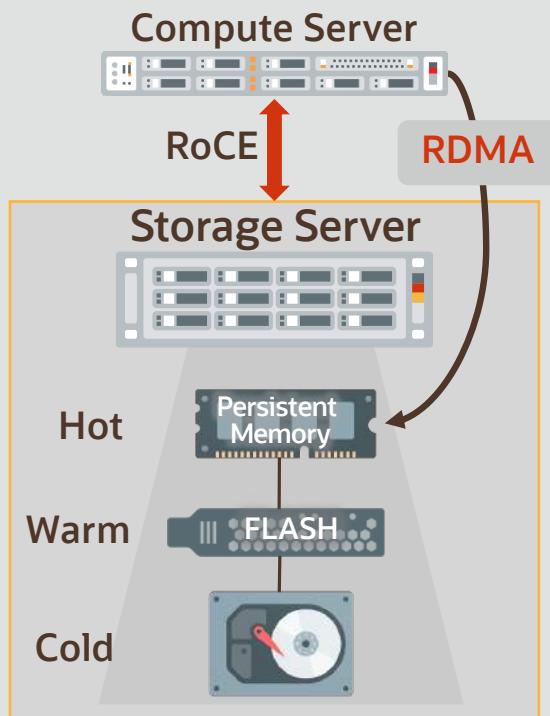
- Exadata Storage Servers transparently add Persistent Memory Accelerator in front of Flash memory
  - 2.5x higher I/Os per second than current – 16 Million IOPS

- Database uses RDMA instead of I/O to read remote PMEM
  - Bypasses network and IO software, interrupts, context switches
  - 10x better latency – under 19 μsec for 8K database read

- PMEM Automatically tiered and shared across DBs
  - Using as a cache for hottest data increases effective capacity 10x

- Persistent Memory mirrored automatically across storage servers for fault-tolerance

Exadata X8M Persistent Memory Commit Accelerator

Compute Server

RoCE

RDMA Log Write

Storage Server

Hot

Persistent Memory

Flush Later to Flash/Disk

Warm

FLASH

Cold

- Log Write latency is critical for OLTP performance
  - Faster log writes means faster commit times
  - Any log write slowdown stalls the whole database

- Automatic Commit Accelerator
  - Database issues one-way RDMA writes to PMEM on multiple Storage Servers
  - Bypasses network and I/O software, interrupts, context switches, etc.
 - Up to 8x faster log writes

PMEM in Shared Exadata Storage Advantages

- **Better utilization of PMEM capacity** – only hottest data moved into PMEM

- **More adaptive** - aggregate performance of PMEM on all storage servers dynamically used by any database on any server

- **End to end security** – PMEM only accessible to databases using database access controls, no OS or local access

- **Simpler** – all PMEM benefits with no extra management

- **More resilient** – automatic mirroring of PMEM across storage servers protects from PMEM failures

- **Much better latency** – RDMA from Database directly to PMEM

Exadata Persistent Memory Delivers Extreme Performance

- Exadata dramatically accelerates most latency-sensitive database operations, reads and commits, using RDMA
  - Database can near-instantly retrieve cached data from storage
  - Database can near-instantly write commit records to storage
- Third-party storage cannot provide RDMA interface between database and storage
  - Unable to unlock true benefit of persistent memory—super-low latency, extreme IOPS
- Exadata X8M provides faster performance unmatched by competing technologies
  - Exadata can consolidate more workloads with the same amount of hardware and software licenses
  - Runs faster with no application changes required
  - Result: Much lower cost database services
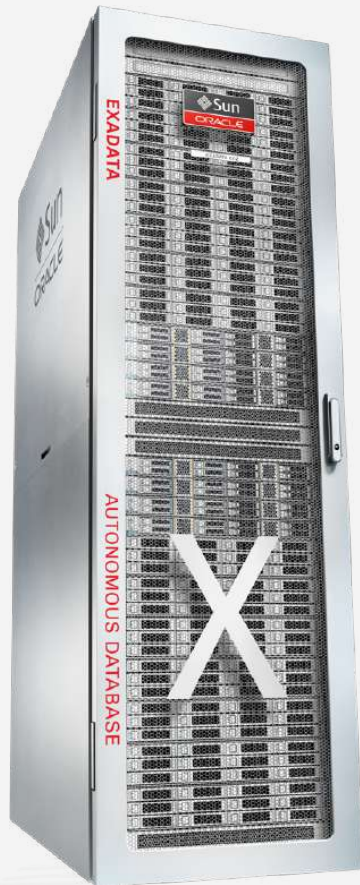
Exadata X8M

Kernel Virtual Machine
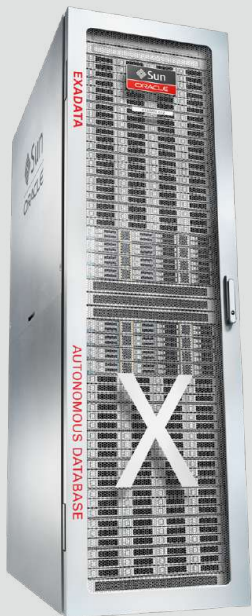
Exadata X8M Virtualization

- Exadata X8M uses KVM-based virtualization
  - Different from Xen-based virtualization used in InfiniBand Exadata
- RoCE and Persistent Memory only supported on KVM

- KVM gives:
  - 2x more guest VM Memory – 1.5 TB/server
  - Faster client network latency
  - 50% more guest VMs per server

Exadata X8M

Customer View

Exadata X8M Benefits

- Get latency of persistent memory, IOPs of flash, cost of disk

- OLTP Workloads
  - 10x lower latency for I/O, 2.5x more I/Os, super fast log writes

- Data-Warehouse and Mixed Workloads
  - PMEM RDMA frees CPU on storage server for more smart scans
  - 100 Gb/sec RoCE gives higher net throughput for operations that need it

- Big Picture
  - Exadata already had highly optimized flash which makes it very fast
  - Biggest PMEM application benefit will be very fast latency for I/O
    - For example faster nested loops joins that perform many I/Os

No Application Changes Required
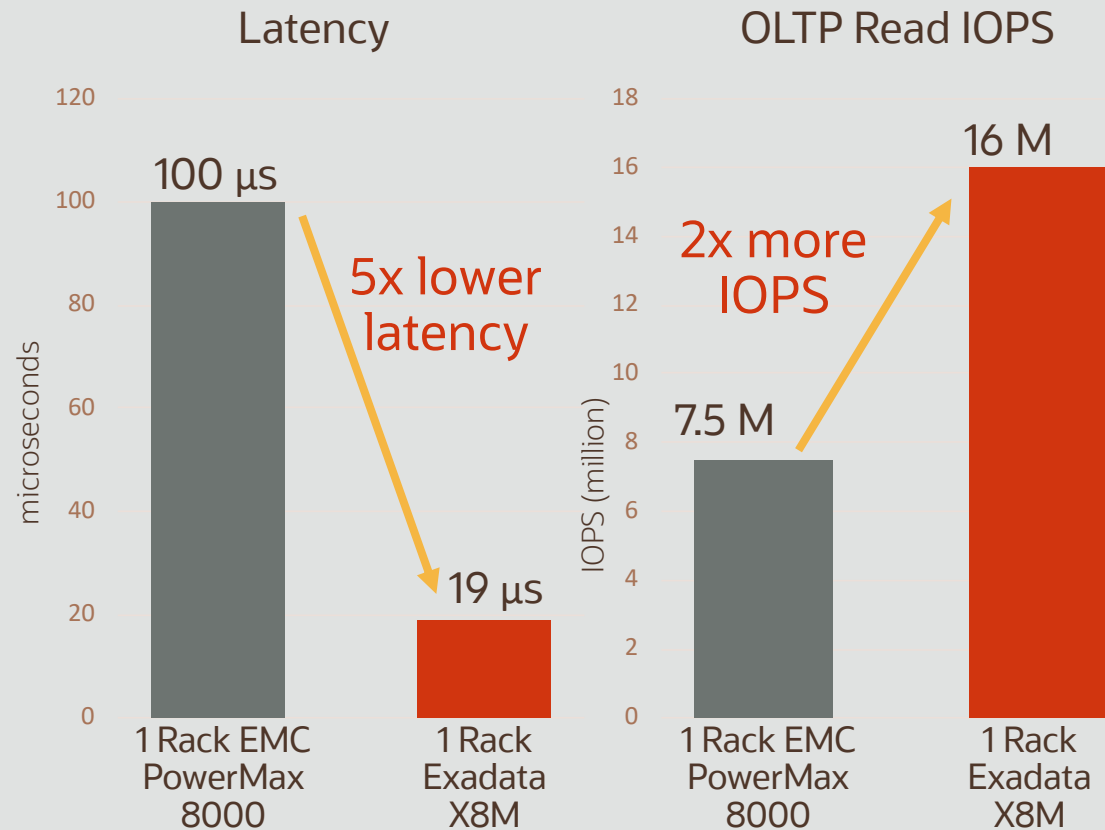
Exadata X8M Business Benefits
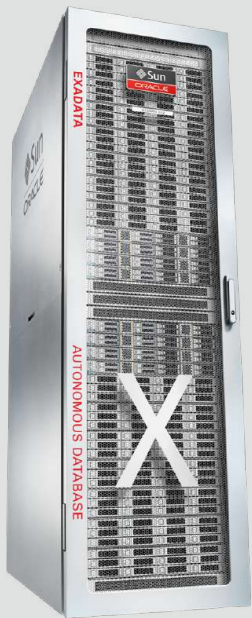
- Faster OLTP and Analytics performance at the same price
  - Perform existing tasks faster and provide knowledge sooner
  - Perform deeper or more frequent analysis
  - Consolidate more workloads on same resources reducing overall costs
  - Eliminate need for costly in-memory workarounds for latency-sensitive workloads
- Performance scales as platform grows
  - Continue to realize performance benefits as workloads grow
  - Elastically expand platform buying capacity as you need it
  - Who benefits
    - Personalized Ads
    - Gaming
    - Fraud Detection
    - Stock Exchanges
    - Telco - Call data records

Exadata X8M I/O is Dramatically Faster than All-Flash EMC

Single rack Exadata X8M beats the fastest EMC PowerMax array in every performance metric

- ✓ 3x more throughput
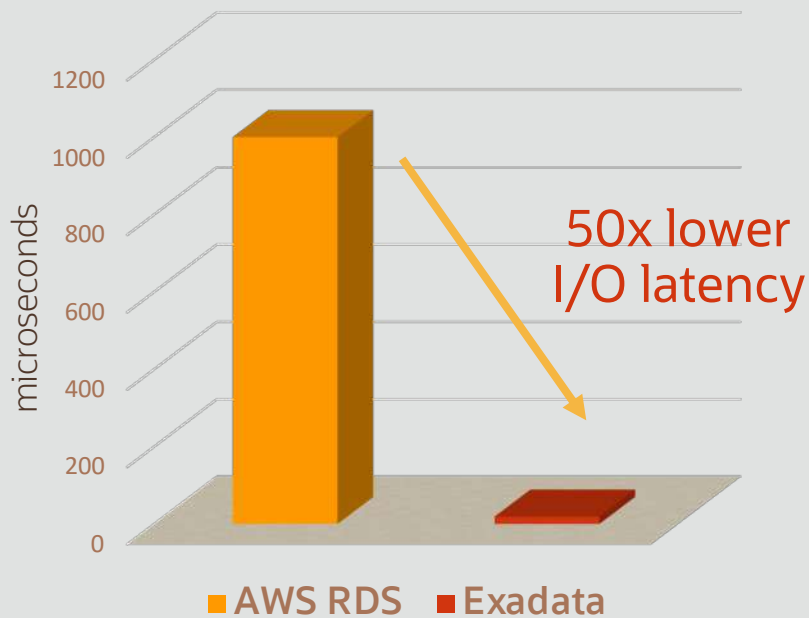- ✓ 2x more IOPS
- ✓ 5x lower latency

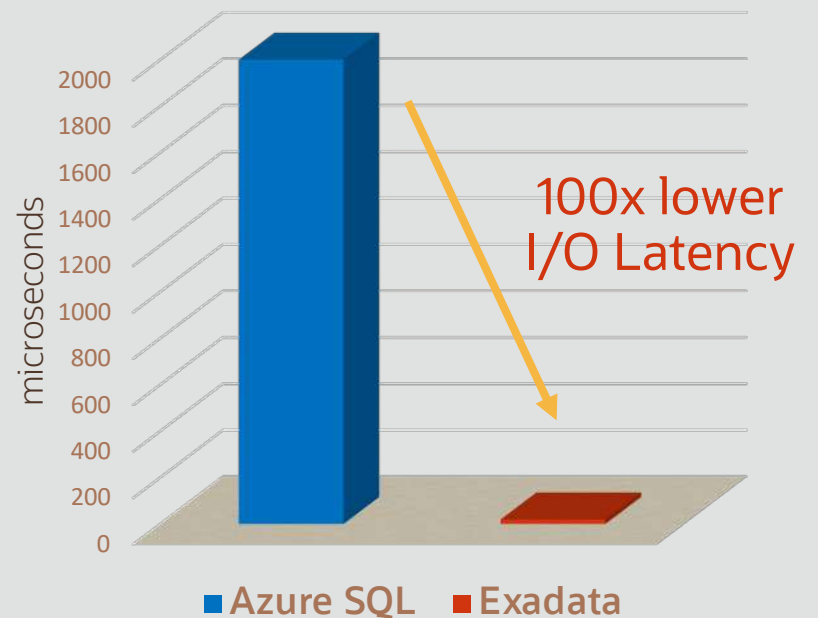Exadata performance scales linearly as more racks are added

Latency

microseconds

100 µs — 1 Rack EMC PowerMax 8000

5x lower latency

19 µs — 1 Rack Exadata X8M

OLTP Read IOPS

IOPS (million)

7.5 M — 1 Rack EMC PowerMax 8000

2x more IOPS

16 M — 1 Rack Exadata X8M

Exadata X8M Storage Performance
>50x faster than Flash Block Storage on AWS or Azure

Flash Storage on AWS vs. Exadata

microseconds

1200
1000
800
600
400
200
0

50x lower I/O latency

■ AWS RDS ■ Exadata

Flash Storage on Azure vs. Exadata

microseconds

2000
1800
1600
1400
1200
1000
800
600
400
200
0

100x lower I/O Latency

■ Azure SQL ■ Exadata

# Exadata X8M Availability

- Exadata version with RoCE, Persistent Memory and KVM is called Exadata X8M

- Exadata X8M on-premises ships Sept 2019

- Exadata X8M promotional price
  - Exadata X8M will initially ship with upgraded ROCE RDMA network and full Persistent Memory for no additional cost
  - 1.5 TB of Persistent Memory per storage server

- Exadata X8 (InfiniBand) will continue to ship
  - For customers expanding existing Exadatas
 - For customers that need to certify X8M before adopting it

ZDLRA X8M

- World's First Engineered System Built for Database Protection
  - Zero Data Loss data protection and recovery validation
  - Efficient incremental forever backup frees production resources
  - Cloud scale enterprise deployment
- ZDLRA X8M uses new 100 Gb/sec RoCE for high throughput internal data transfers between compute and storage
  - Network to clients unchanged: 25 Gb/sec Ethernet
  - ZDLRA X8M on-premises ships Sept 2019 - no price increase
- ZDLRA X8 (InfiniBand) will continue to ship for customers expanding existing ZDLRAs
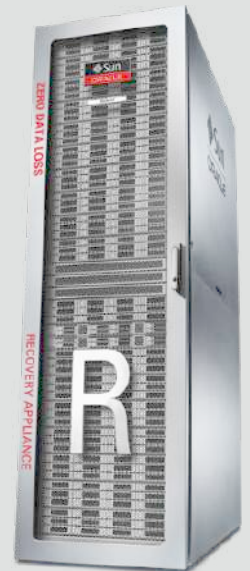
Oracle's Commitment to Exadata is Increasing
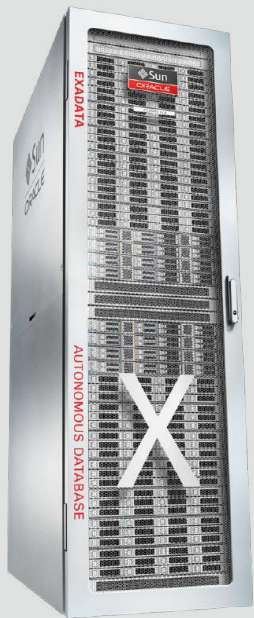
- Oracle SaaS apps (ERP, HCM, etc.) run exclusively on Exadata
  - Thousands of Exadata systems deployed globally in Oracle public cloud

- Autonomous Database services run exclusively on Exadata

- Exadata Cloud at Customer brings full cloud benefits to customer data center

- Exadata on-premises remains a major focus
  - Will exclusively run all **database features** developed for Autonomous DB
  - Even better platform for private database cloud

Exadata Advantages Increase Every Year

Dramatically Better Performance and Cost

• KVM Virtualization
• Autonomous Database
• Automatic Indexing
• Exadata Cloud at Customer
• In-Memory OLTP Acceleration
• In-Memory Columnar in Flash
• Exadata Cloud Service
• Smart Fusion Block Transfer

• RoCE Networking
• Persistent Memory

• In-Memory Fault Tolerance
• Direct-to-wire Protocol
• JSON and XML offload
• Instant failure detection

• Hot Swappable Flash
• 25 GigE Client Network

• Network Resource Management
• Multitenant Aware Resource Mgmt
• Prioritized File Recovery

• 3D V-NAND Flash
• Software-in-Silicon

Smart Software

• IO Priorities
• Data Mining Offload
• Offload Decrypt on Scans

• Tiered Disk/ Flash

• Database Aware Flash Cache
• Storage Indexes
• Columnar Compression

• PCIe NVMe Flash

• Unified InfiniBand

• Smart Scan
• InfiniBand Scale-Out

• DB Processors in Storage

Smart Hardware

• Scale-Out Storage

• Scale-Out Servers

Oracle OpenWorld 2019