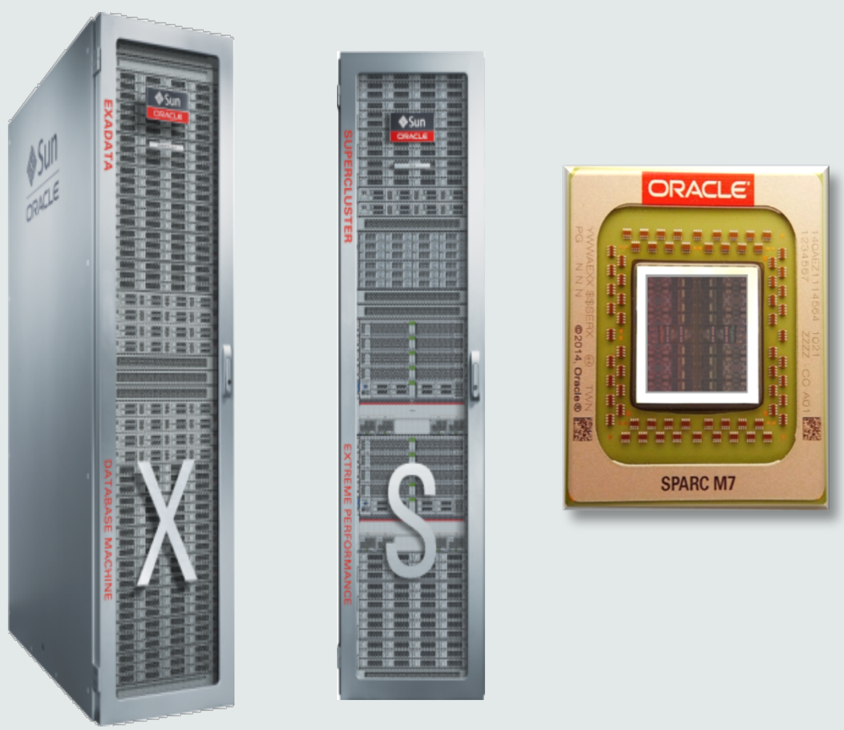
Engineering Database Hardware and Software Together

From Engineered Systems to SQL in Silicon



Juan Loaiza
Senior Vice President
Oracle

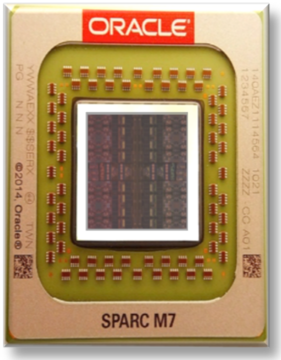
ORACLE

Safe Harbor Statement

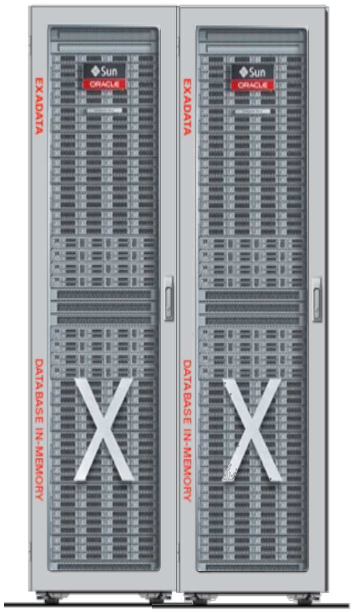
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development,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

Engineering Database Hardware and Software

- Existing Engineered Systems deeply integrate Database Software with best-of-breed hardware
  - Exadata and Supercluster
- This year Oracle extends integration to the Microprocessor Level
- Customer Benefits:
  - Performance, Reliability, Cost, Security

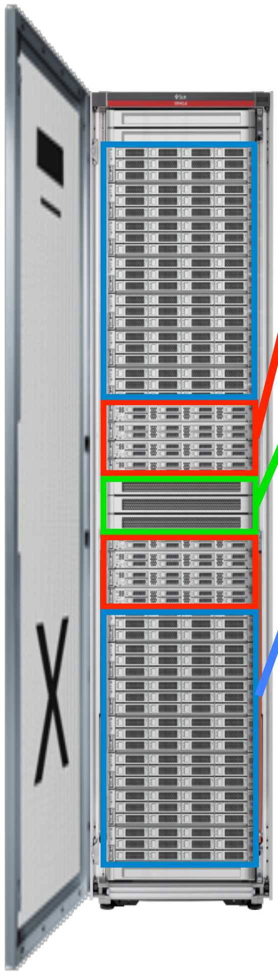


Oracle Exadata Database Machine



- Scale-out, database optimized compute, networking, and storage hardware for fastest performance and lowest costs
- Unique software and protocols enable fastest and most efficient OLTP, Analytics, and Consolidation
- Delivered integrated, optimized, automated, and supported end-to-end to reduce operations costs

Exadata Hardware



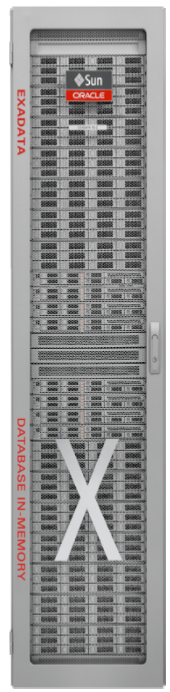
- **Compute Servers**
  - Latest fastest processors, largest memory
- **Unified Network -** InfiniBand for fastest performance
- **Storage Servers**
  - 2-socket servers - low-cost and power CPUs
  - Fastest PCIe flash combined with high capacity disk performance, capacity, and cost
  - Data is duplicated across storage servers for high availability
- **Scale by adding more compute or storage servers as needed**

Compute Server

Storage Server

System Level Engineering of Database Software and Hardware

- **Data Warehousing**
  - SQL Offload to storage servers enables full flash bandwidth (100s of GB/sec)
  - Flash bandwidth is too high for network
- **OLTP**
  - Database calls InfiniBand NICs directly bypassing O/S to achieve millions of IOPS
  - Smart Flash logging accelerates commits
- **Availability**
  - Instant server death detection by integrating with InfiniBand switches
  - Sub-second I/O failover caps I/O latency
  - Mirroring of In-Memory Database data



- **Storage**
  - Achieve flash speed with disk capacity by intelligently caching data in flash
  - Flash Cache automatically converts data to columnar format for faster analytics
- **Safe Consolidation**
  - I/Os prioritized in storage servers by database, or SQL user, or SQL job
  - Low latency network traffic uses separate network lanes from high bandwidth traffic
    - E.g. separate Commit message and Report

Thousands of Mission Critical Deployments

- **Petabyte Warehouses**
- **Business Applications**
  - SAP, Oracle, Siebel, PSFT, …
- **Online Financial Trading**
- **E-Commerce Sites**
- **Massive DB Consolidation**
- **Public SaaS Clouds**
  - Oracle Fusion Apps, NetSuite, Salesforce, …

4 out of the 5 Largest Banks, Telecoms, Retailers Run Exadata

ORACLE

Deployments are Mix of OLTP and Analytics

- Half of deployments primarily OLTP, half Analytics
  - Many run mixed workloads
- Some say you need a specialized product for each workload, Oracle and the market disagree
  - Key is specialized algorithms, not specialized products
  - Specialty products die as specialized algorithms are added to general databases
- General databases have 4 big advantages
  - Handle mixed and complex use cases
  - Less need for complex cross product data motion
  - Much better operational attributes
    - Security, management, backup, availability, scaling, etc.
  - Simpler – less moving parts to learn/operate/patch/secure

ORACLE

Next Big Integration Focus: In-Memory Database

Oracle Database in Memory DB, Released Summer 2014

Real Time Analytics

Directly on OLTP Data

No Changes to Applications

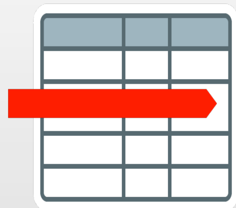
Trivial to Implement

ORACLE

Row Format Databases vs. Column Format Databases

Rows Stored Contiguously

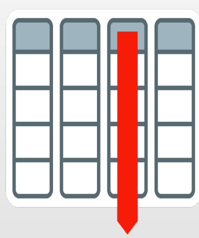
SALES



- **Transactions** run faster on row format
  - Example: Query or Insert a sales order
  - Fast processing few rows, many columns

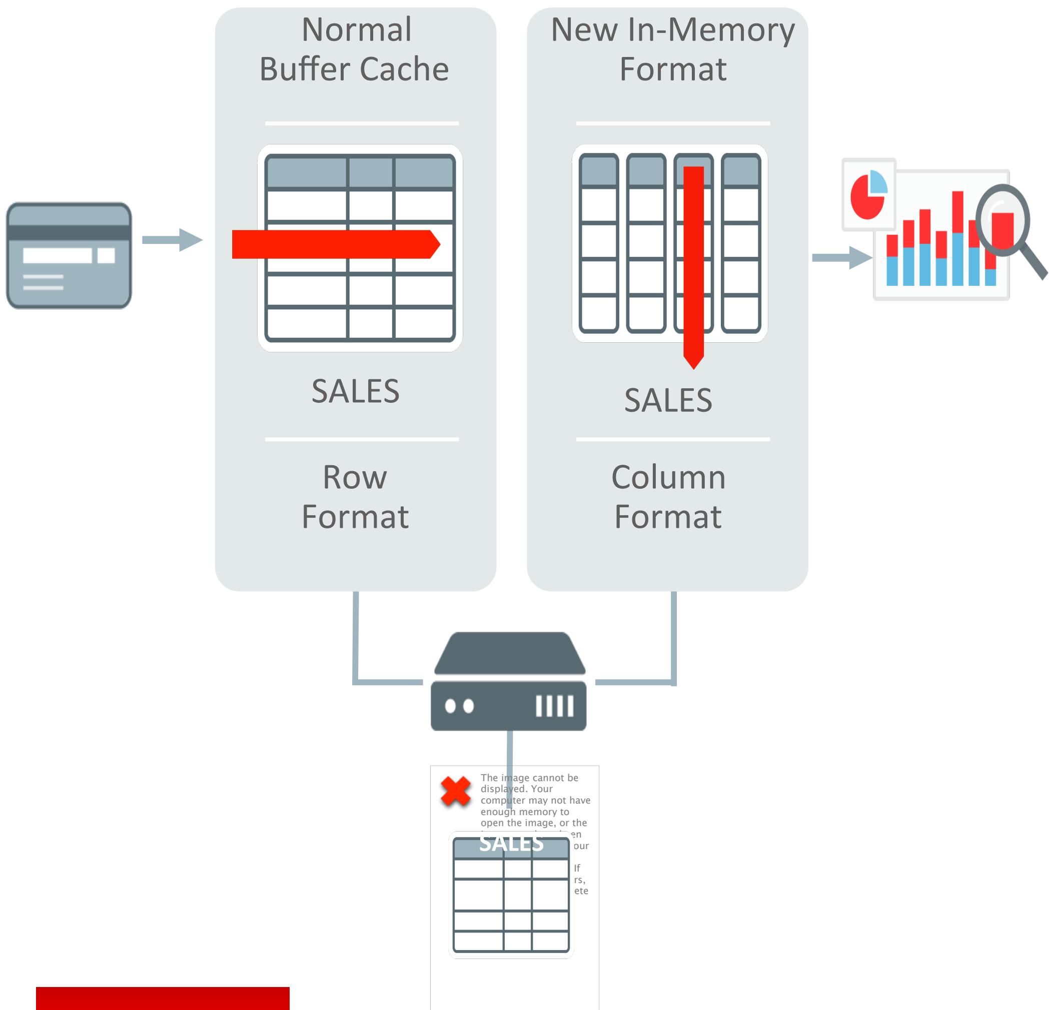
Columns Stored Contiguously

SALES



- **Analytics** run faster on column format
  - Example : Report on sales totals by region
  - Fast accessing few columns, many rows

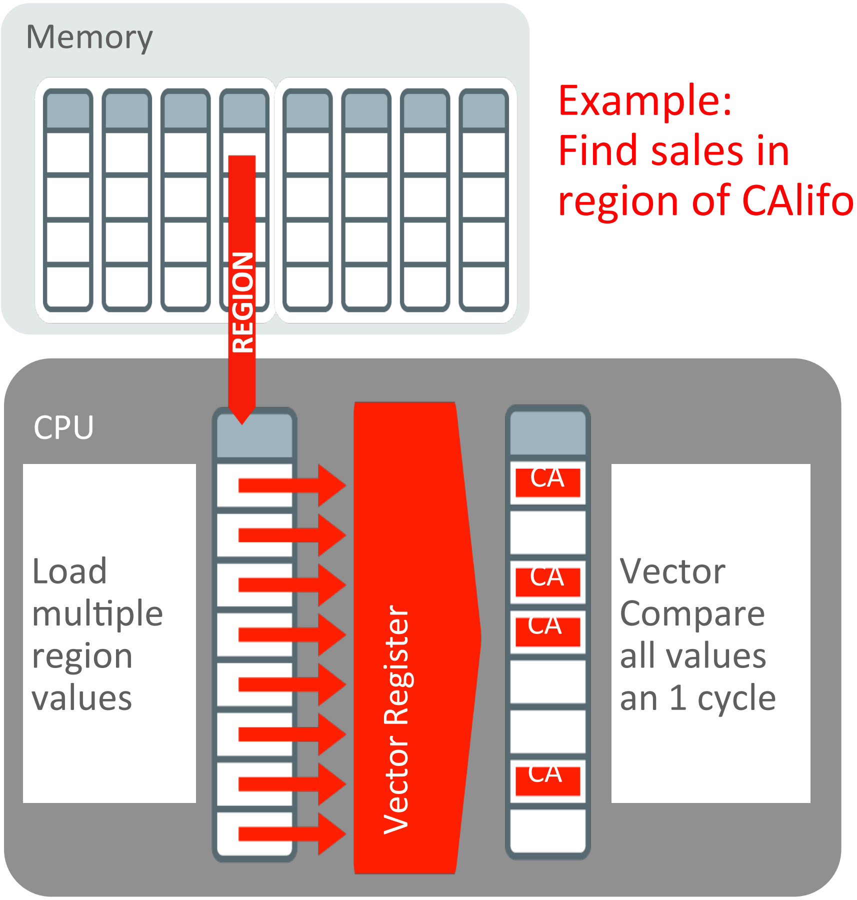
Oracle Dual Format Architecture



- **BOTH** row and column formats for same table
- Simultaneously active and transactionally consistent
- OLTP uses proven row format
- Analytics & reporting use new in-memory Column format
  - Not persistent, and no logging
  - Quick to change data: **fast OLTP**
- Full Scale-Out and Scale-Up

ORACLE

Orders of Magnitude Faster Analytic Data Scans



Example:
Find sales in region of CAlifornia

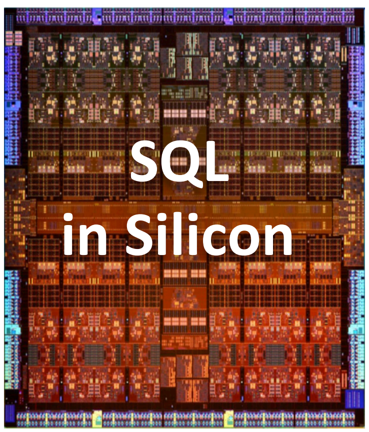
> 100x Faster

- Each CPU core scans local in-memory columns
- Scans use fast SIMD vector instructions
- **Billions of rows/sec** scan rate per CPU core
  - Row format is millions/sec

ORACLE

Coming in 2015: SPARC M7 SQL in Silicon

SPARCing a SQL Performance Revolution



- Traditional DB algorithms too complex for chips
  - Code is large with lots of branches and random accesses
  - Speedups came from more CPU cores, bigger caches, etc.
- Big Change: In-memory columnar is much simpler
- 5 years ago Oracle initiated a revolutionary project
  - Build fastest ever conventional microprocessor

ORACLE

In-Memory Algorithms Natively Implemented on Silicon

Performance

DB In-Memory Acceleration Engines

Reliability/Security

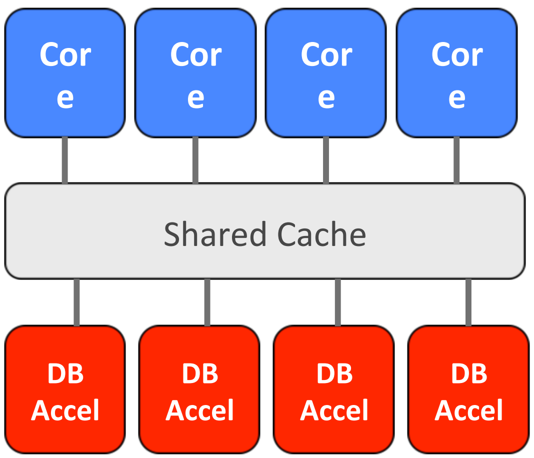
Application Data Integrity

Capacity

Decompression Engines

Performance: Database In-Memory Acceleration Engines

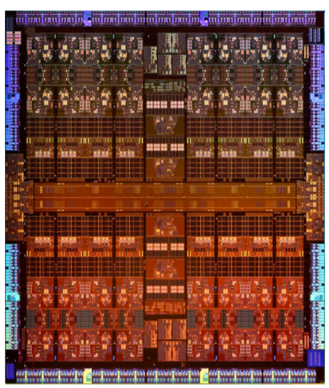
SPARC M7



32 Database Accelerators (DAX)

- SIMD Vectors instructions were designed for graphics, not database
- New SPARC M7 chip has 32 optimized database acceleration engines (DAX) built on chip
- Independently process streams of columns
  - E.g. find all values that match 'California'
  - Up to 170 Billion rows per second!
- Like adding 32 additional specialized cores to chip
  - Using less than 1% of chip space

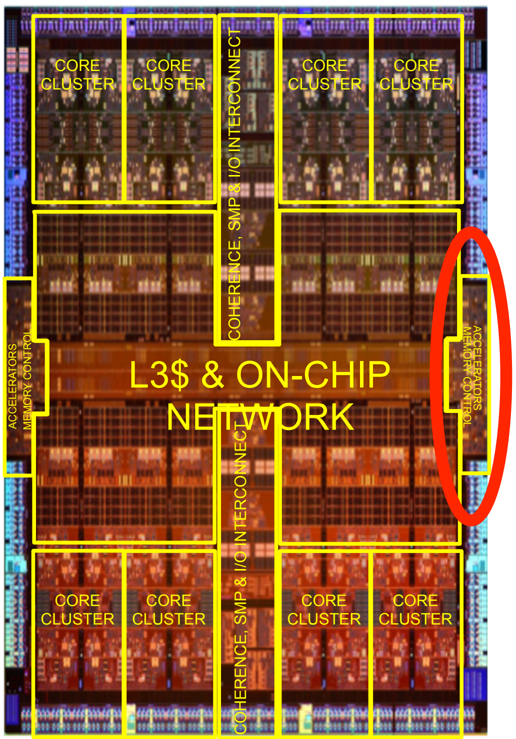
Capacity: Decompression Engines

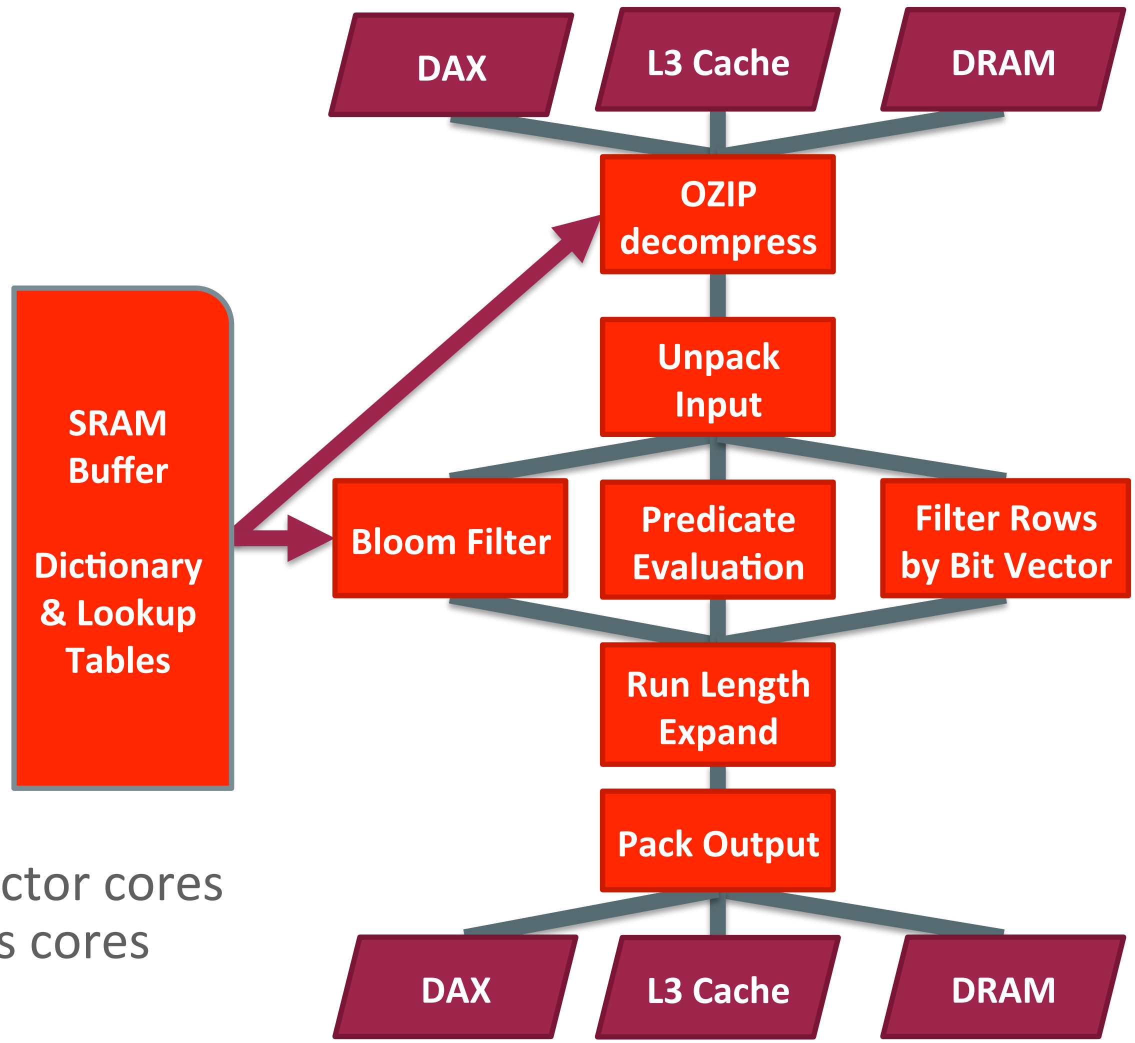


Doubles Memory Capacity

- Compression is key to putting more data in-memory
  - Databases compress repeated symbols, e.g. repeat of ‘California’
  - Don’t compress bit patterns - letter ‘e’ more common than ‘z’
  - Bit pattern compression gives approximately 2X more capacity
- Decompression is far more import for databases than compression
- Bit pattern decompression in normal cores is slow
  - Performance of decompress on today’s processors is fine for disk data, slow for flash, huge bottleneck for in-memory database
  - 64 CPU cores needed to decompress at full memory speed
- SPARC M7 adds 32 decompress engines
  - Run bit-pattern decompress at memory speed

Database Accelerators (DAX): Pipelined Streaming Engines

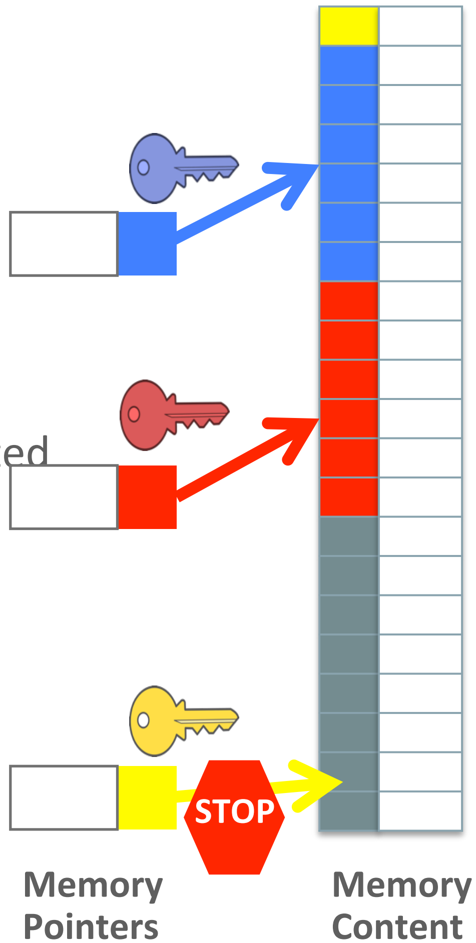




Equivalent of 32 extra vector cores
plus 64 extra decompress cores

Reliability & Security: Application Data Integrity

- Database In-memory places terabytes of data in memory
  - More vulnerable to corruption by bugs/attacks than storage
- SPARC M7 Application Data Integrity implements fine grained memory protection with negligible impact on performance
- Hidden "color" bits added to pointers (key), and content (lock)
- Pointer color (key) must match content color or program is aborted
  - Set on memory allocation, changed on memory free
- Helps prevent access off end of structure, stale pointer access, malicious attacks, etc. plus improves developer productivity

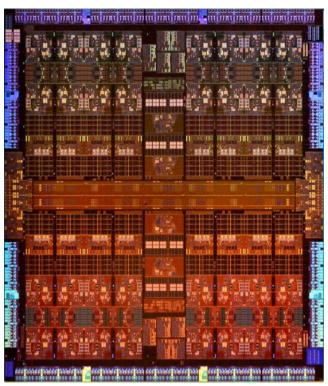


ORACLE

How Data Protection Works

64-bit Pointer

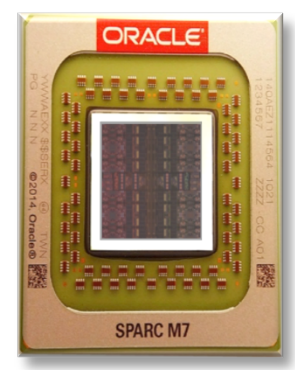
Color Bits	Virtual Address



- Color bits kept in upper bits of pointers
- Color bits kept for every 64 byte aligned memory region
  - Similar to ECC (Error Correction) bits but designed to protect data from for software failures not hardware
  - 2 to 5 orders of magnitude finer granularity than OS pages
- Color bits present in entire memory architecture:
  - All memory paths
  - Full Cache Hierarchy
  - All Chip Interconnects
  - Color bits checked by core load/store units

Conclusion

- Engineering Database Hardware and Software
  - Faster, More Reliable, More Cost Effective, More Secure databases
- Oracle Exadata is a Hardware Platform integrated with the Database at the System Level
- This year Oracle Sparc M7 extends database integration to the Microprocessor – SQL in Silicon
- The beginning of a new era of Database integration
  - **Many opportunities for future research and algorithms**



ORACLE

ORACLE®