Your. Open. World.

Oracle Database 11g: Next Generation Performance and Scalability

Juan Loaiza
Senior Vice President

ORACLE

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.

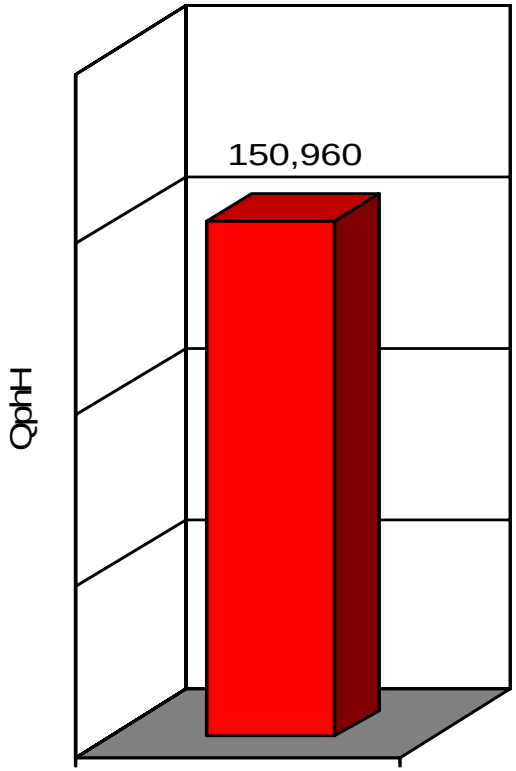
Oracle is Ready to Scale

Leader in Industry Benchmarks

Benchmark	World Record Leadership
TPC-C Clustered	Oracle
TPC-C Price/Performance	Oracle
TPC-H @ 1,000 GB Non-Clustered	Oracle
TPC-H @ 3,000 GB Non-Clustered	Oracle
TPC-H @ 10,000GB Non-Clustered	Oracle
TPC-H @ 30,000 GB	Oracle
SAP Sales and Distribution Parallel	Oracle
SAP Sales and Distribution 2-tier	Oracle
SAP (ATO) Assemble-To-Order 2 and 3 Tier	Oracle

As of July 17, 2008: Source: www.tpc.org & www.sap.com/benchmark SAP TRBK Standard Application Benchmark: Sun Fire E6900 DB Server (8 1.5 GHz US-IV+ processors, 16 cores, 16 threads, 56 GB memory): 10,012,000 Day posts/hr, 6,664,000 Night bal accs/hr, Solaris 10, Oracle 10g. SAP Account Management 3.0 (64-bit) Cert #2006018.The two-tier SAP Business Information Warehouse 3.5 Standard Application Benchmark suite performed on 2/28/06, by Fujitsu Siemens Computers in Paderborn, Germany, was certified on 3/14/06 with the following data. The scenario for 32GB main memory which corresponds to 467,200,000 records in fact table was used. Load Phase - Average throughput total step 1+2 (rows/hour): 53,255,652. Query Navigation Steps: 377,280. The software configuration for all steps of the SAP BW Benchmark: Operating system central server: SUN Solaris 10. RDBMS: Oracle 10*g*. Platform Release: SAP NetWeaver '04. Configuration: Central server: Fujitsu Primepower 850, 16 processors/16 Cores/16 threads, SPARC64 V, 2.16 GHz, 128 KB (D) + 128 KB(I) L1 cache, 4 MB L2 cache, and 32 GB main memory, Cert # 2006014.

Best Data Warehouse Performance - World Record 30 TB TPC-H



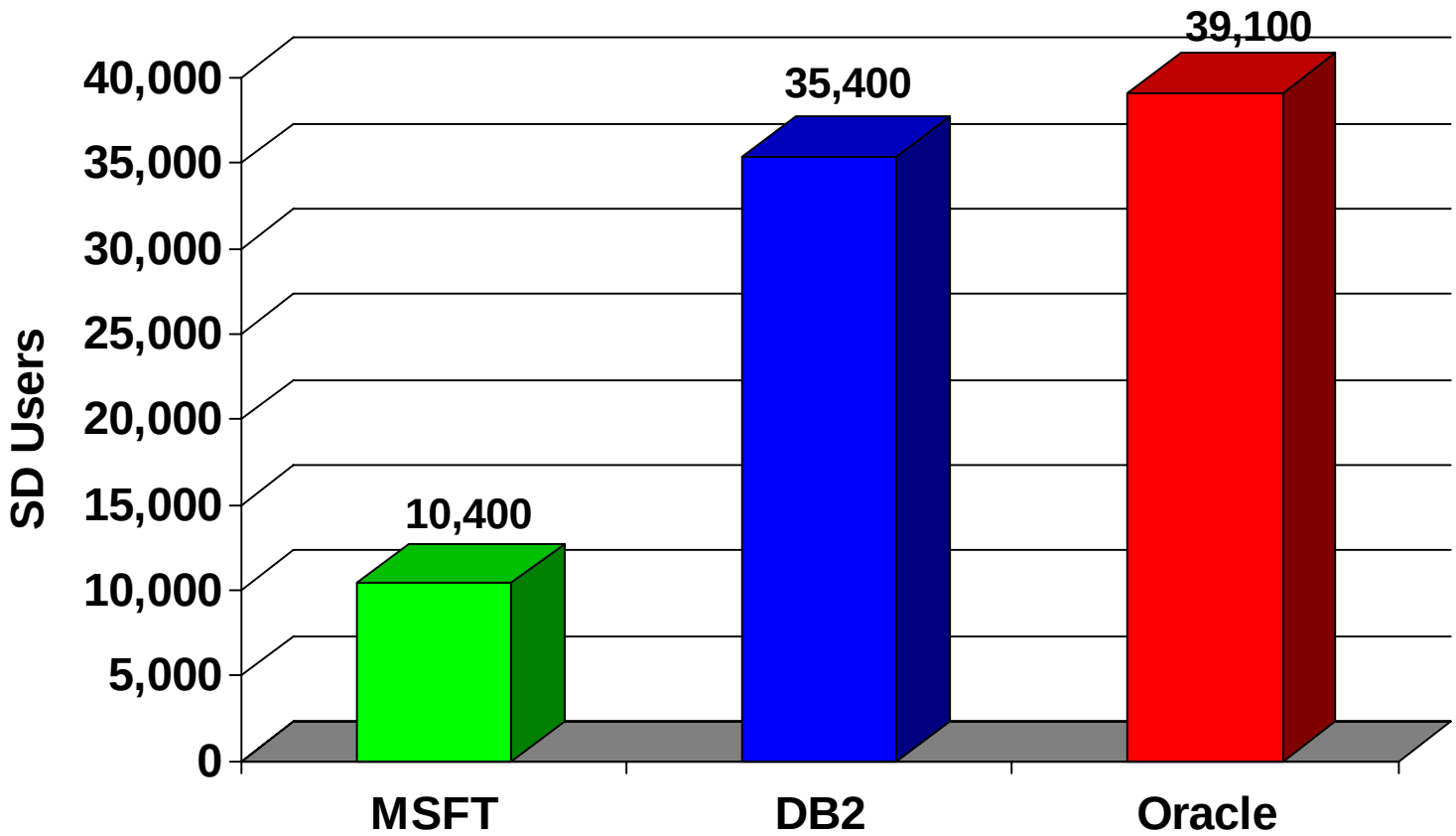
Processor Cores	**128**
Memory	**1 Terabytes**
Disk Arrays	**256 MSA1000's**
Disk Storage	**448TB**
Read Throughput	**40GB/Sec**

Source: TPC, As of Nov 9, 2007: Oracle Database 10g on HP Superdome Server, 150,960 QphH $46.69/QphH, avail 6/18/07.

Best Business Performance

- **World Record SAP SD 2-tier Benchmark**

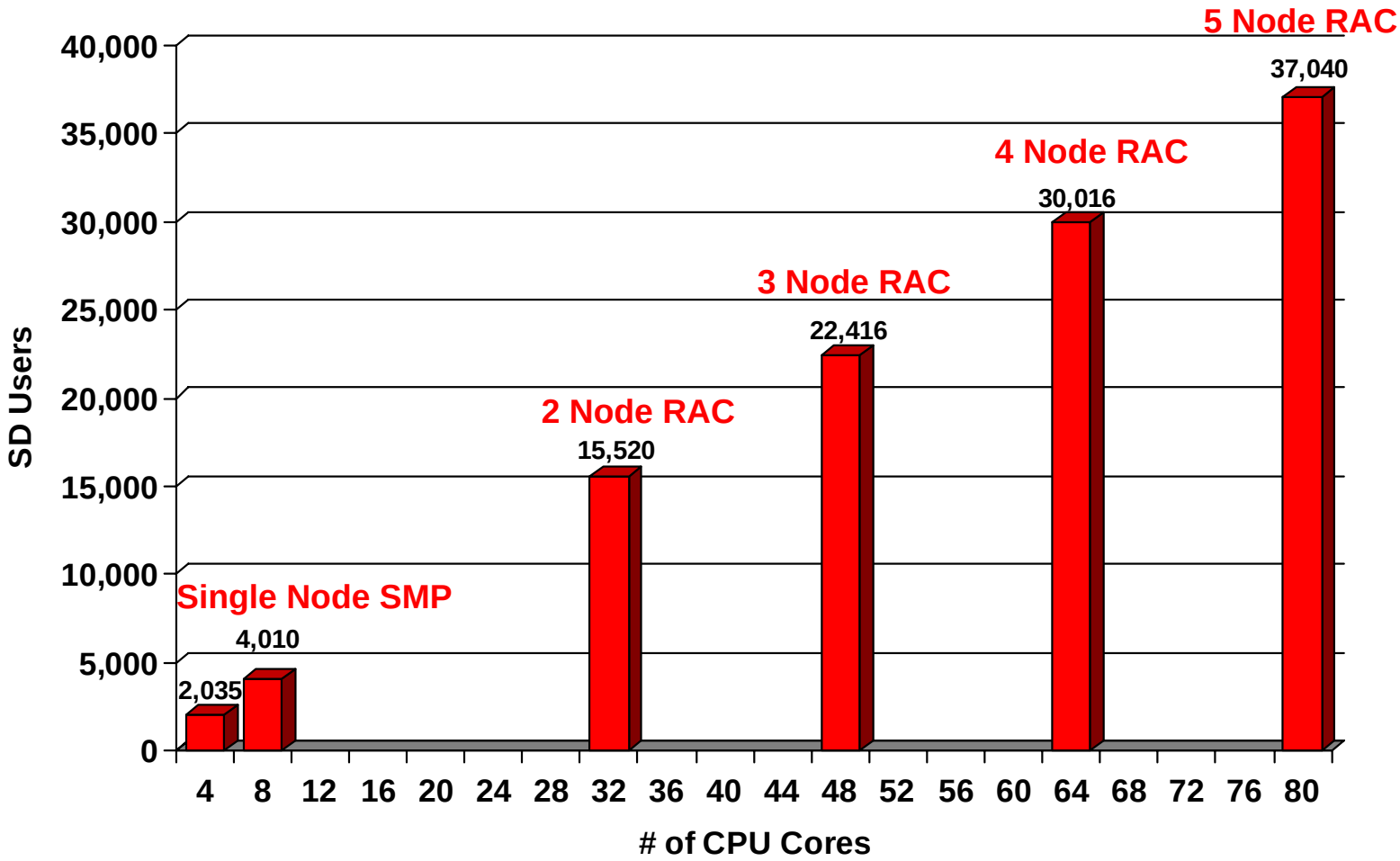
First Ever Benchmark on a 256 Core SMP



These results, as of July 14, 2008, have been certified by SAP AG, www.sap.com/benchmark. Please see notes page for benchmark certification details for the above results.

Best Scalability and Performance

World Record SAP SD Benchmark Results

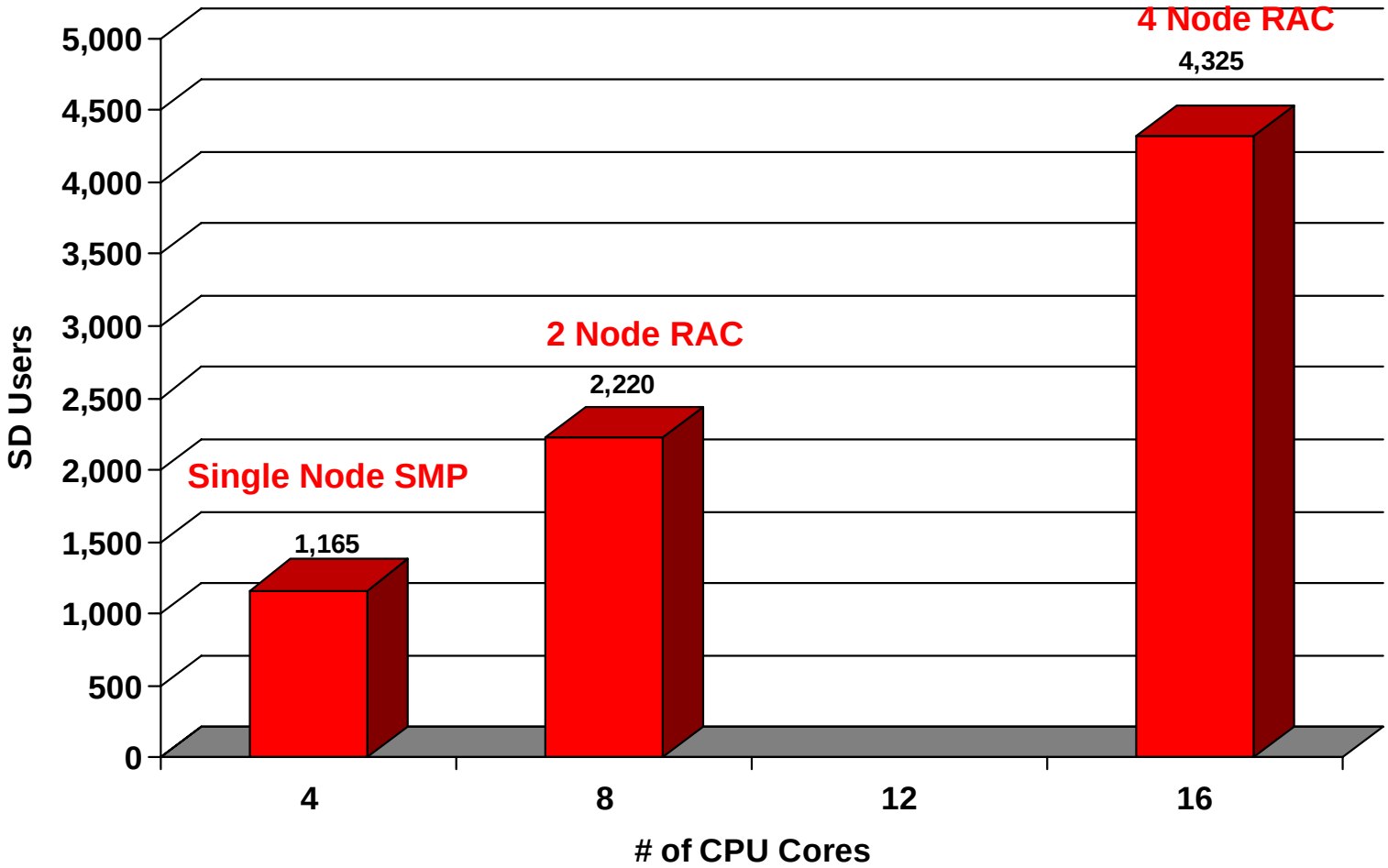


Near Perfect Scaling across SMP and Cluster

These results, as of March 25, 2008, have been certified by SAP AG, www.sap.com/benchmark. Please see notes page for benchmark certification details for the above results.

Best Scalability and Performance

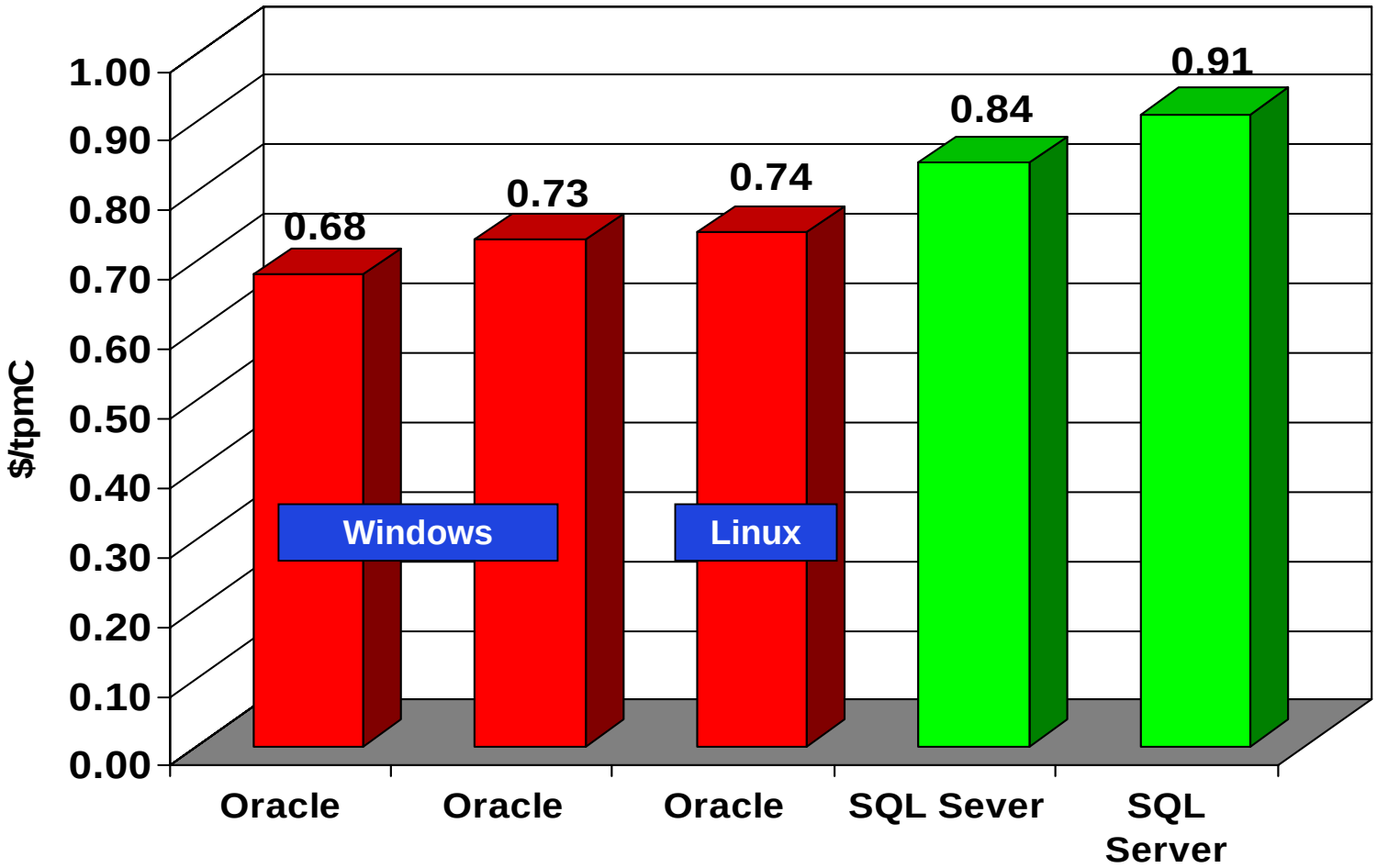
World Record SAP SD Benchmark Results



Near Perfect Scaling across SMP and Cluster

These results, as of Sep 22, 2008, have been certified by SAP AG, www.sap.com/benchmark. Please see notes page for benchmark certification details for the above results.

Oracle Database Holds Top 3 TPC-C Price Performance



Unprecedented Value Leadership Over Microsoft

As of June 24 2008: Dell PowerEdge 2900 97,083 tpmC, .68/tpmC, available 6/16/08. HP ProLiant ML350G5, 102,454 tpmC, .73/tpmC, available 12/31/07. HP ProLiant ML350G5, 100,926 tpmC, .74/tpmC, available 6/8/07. Microsoft SQL Server on HP ProLiant ML350G5, 82,774, .84/tpmC, available 03/27/07. Dell PowerEdge 2900, 69,564 tpmC .91/tpmC, available 3/9/07. Source: Transaction Processing Performance Council (TPC), www.tpc.org

Leader in Real-World Scalability

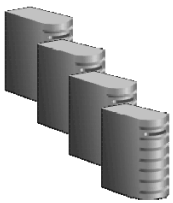
- Winter TopTen™- survey of largest and most heavily used databases in the world
  - Sizes are table sizes after compression
    - RAID, mirroring and free space NOT counted
- The World's Largest Commercial Database
  - Yahoo @ 100TB
- The World's Largest Linux Data Warehouse
  - Amazon.com @ 24.7TB
- The World's Largest Linux OLTP systems
  - Amazon.com @ 4TB
- The World's Largest Scientific database
  - Max Planck @ 222TB
- The World's Largest Unix OLTP database
  - US Patent & Trademark Office @ 16.4TB
- Oracle runs nine of the ten largest UNIX OLTP systems
- Oracle runs 100 Percent of all Linux DSS and OLTP systems

ORACLE

Source: WinterCorp 2005 TopTen™ Program, www.wintercorp.com

Oracle is Ready to Scale Today

- 25 Years of Investment



- **Scalable Execution**
  - Scale-up on Large SMPs
  - Scale-out on Clusters
  - Parallel Everything
    - Query, DML, DDL, Loads, Unloads, Recovery
  - Non-escalating Row Locks
  - Read Consistency

- **Scalable Availability**
  - RAC
  - Flashback
  - Data Guard
  - RMAN
  - Online Table Redefinition
  - Read-only Tablespaces

- **Scalable Storage**
  - Table Partitioning
  - Automatic Storage Management
  - Bigfile Tablespaces
  - Transportable Tablespaces
  - Nine Index Types

- **Scalable Management**
  - Grid Control
  - Automatic Tuning of Highly Complex Applications
  - Database Resource Manager
  - Automatic Workload Repository

SMP Scale-Up

- Very mature
  - 20 years of experience
- Many customers with largest SMPs on the market
  - 64 to 128 CPUs
  - Sun E25K, HP Superdome, IBM Regatta
- Single System Image
  - Easy to manage
  - Easy to write applications to
- Works great, but high cost
  - Eventually hits a wall
- Need at least one more for Availability

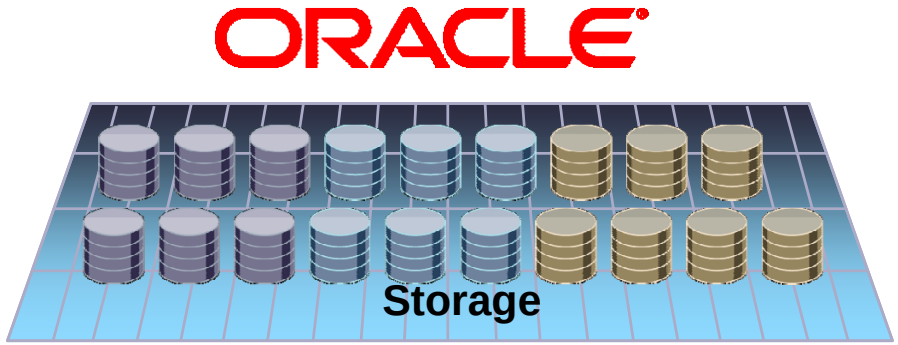
RAC Scale-Out

- RAC pools multiple servers into one virtual server
- Great for Scalability and Availability
- Integrated clusterware
  - Failover to any other node
  - Application notification
- No Idle Resources
- Single System Image
- Thousands of production customers

RAC Scalability

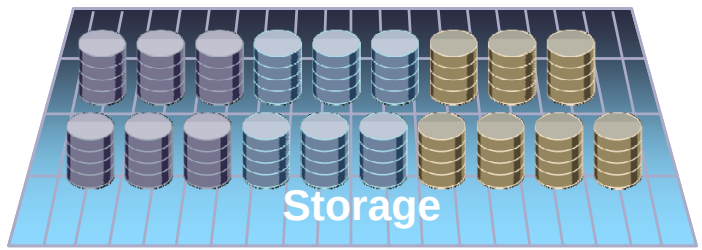
- RAC has no inherent scalability limits
- Very low network usage for OLTP
  - Dozens of processors on 1 gigabit ethernet
- Scale limits are application related, not architectural
  - e.g. all sessions update same row
  - The same limits apply to SMP or any other architecture

Grid Based Storage Scaling



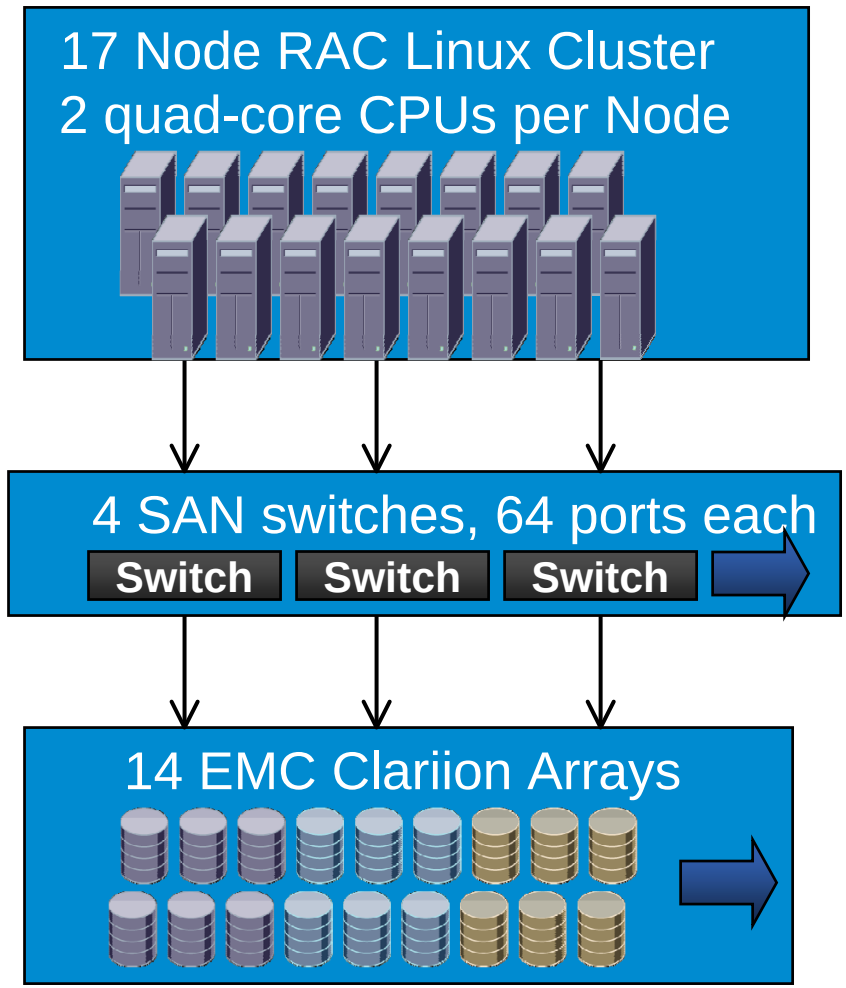
- Grid technology for database storage is emerging
- Oracle's vision is to create a database storage grid using a pool of networked low cost storage blades
- Oracle software will provide the grid attributes
  - Makes it look to users like a single system
  - Manage like a single system
  - Scale and load balance automatically
  - Transparently tolerate component failures

High Scale Examples

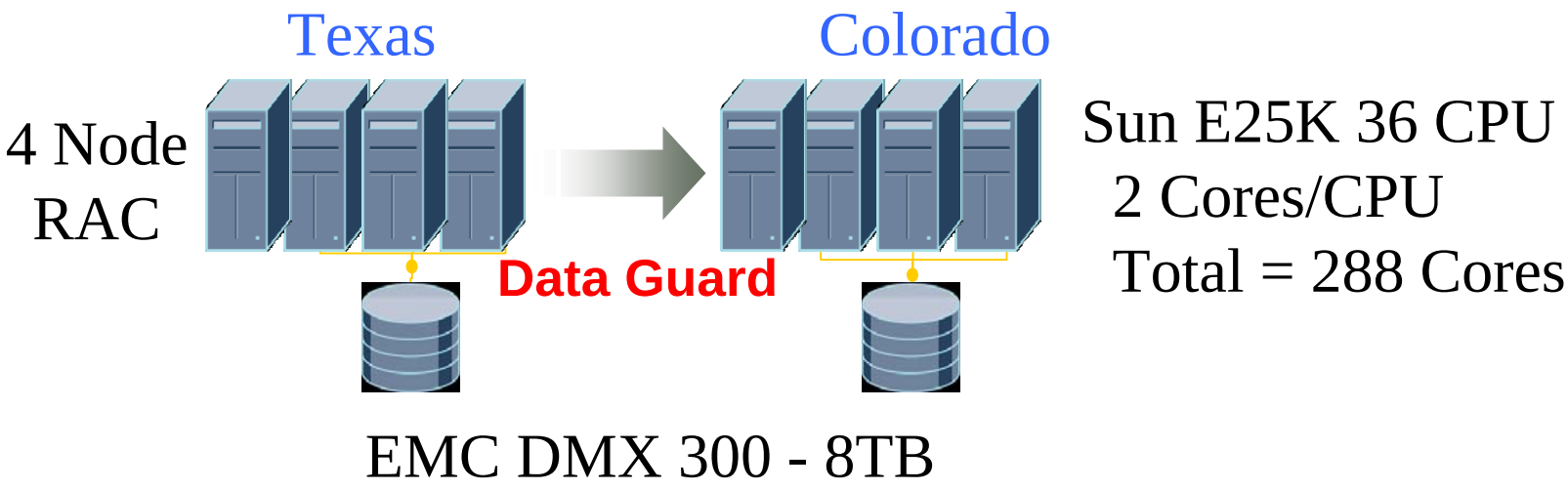


Example: Amazon Data Warehouse

- 100+ TB row data
  - >30 TB of which is compressed 30-40%
  - 320 TB of disk space
- Runs 3 of these configs for DR
  - Each 100+ TB cluster is in geographically separate location
- RAC and ASM
  - Scale-Out architecture
  - Both server and storage layer
- In top ten largest Warehouses in Winter Survey
- 100x data growth in 7 years
- Still seeing 30-40%+ growth each year even using Compression

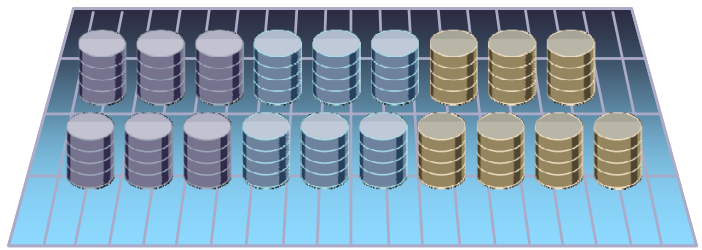


Example: Oracle Central e-Business DB



- Worldwide Central E-business database for Fortune 200 company
- ERP, HR, and CRM
  - Payroll, contracts, procurement, expense reports, hiring, etc.
- Consolidated 70 separate Applications databases
  - Estimated cost savings of over $1B

The Need for More Speed



Every Business Needs Bigger Databases

- More business applications
- Web increases touch points with customers
- Self-service applications for employees
- Data retention regulations
- Multimedia explodes data sizes
- Mergers
- Consolidation of data into fewer databases
- Documents move to XML and into databases

Upcoming Scaling Milestones

- By 2010 we expect to have:

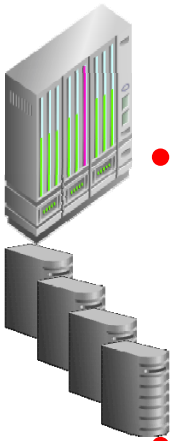
- First Petabyte (1000 TB) Database
  - Biggest sites today around 200TB
  - LOB data is growing particularly quickly
- First 1000 Processor Core Database
  - Biggest sites today in the 100 to 300 processor core range
- First Terabyte buffer cache in production
  - We routinely run with these in benchmarks

The New Normal

- Hardware capacity increases exponentially
  - Megabytes per disk
  - Processor cores per chip
  - Megabytes per DRAM
- Makes massive scale systems affordable to the mainstream
- Terabyte databases were myths ten years ago, now they are common
- In ten years 100TB databases with 1000 Cores will be common
  - 11g will still be in use

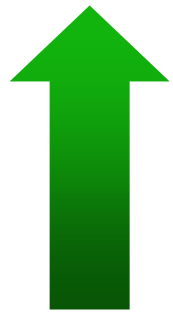
DBMS'es Must Keep Scaling

- Databases don't get a free ride on hardware trends
- Supporting the new normal with a thousand cores and petabytes of data requires new algorithms



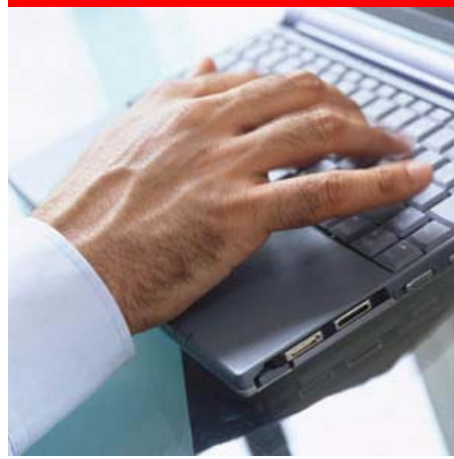
- **Scalable Execution**
  - Must scale to 1000 Cores
    - Parallelism, locks, connections
- **Scalable Storage**
  - Disks get bigger but not faster
  - Compression to speed data access
  - Faster LOBs
  - More sophisticated data partitioning

- **Scalable Availability**
  - Scale Backup and recovery to Petabyte databases
  - Larger Clusters
  - Use Standby to improve scaling



- **Scalable Management**
  - Tuning and management of larger systems, more users
  - Tuning and management of ultra-complex applications
    - e.g. Oracle Fusion Apps

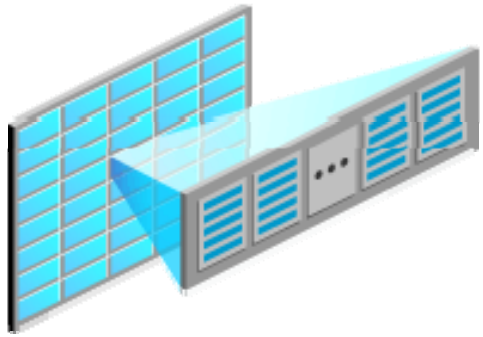
Oracle Database 11g Innovations



- Scalable Execution
- Scalable Storage
- Scalable Availability
- Scalable Management

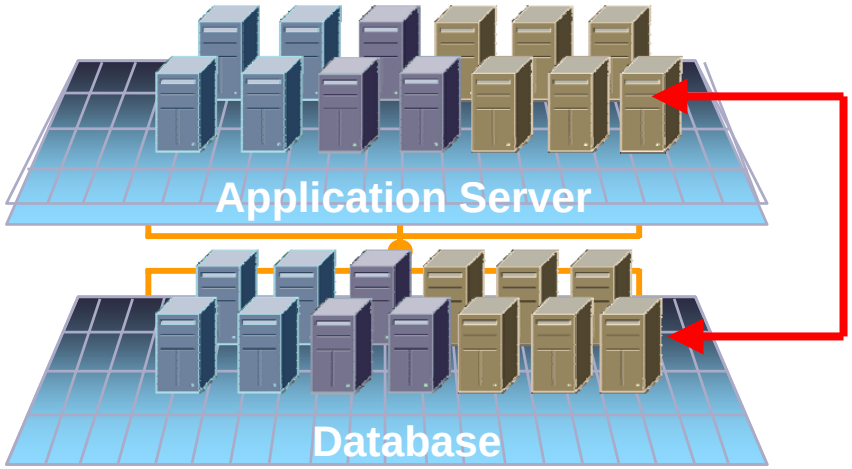
Scalable Execution

Server Results Cache



- Database caches results of queries, sub-queries, or pl/sql function calls
  - Cache is shared across statements and sessions on server
  - Full consistency and proper semantics
- 2x speedup for worst case of trivial query
- 100x speedup for complex queries
- Statement hints specify caching - */*+ *result_cache* +*/*
- Only for very read intensive tables

OCI Consistent Client Cache



Consistent Caching

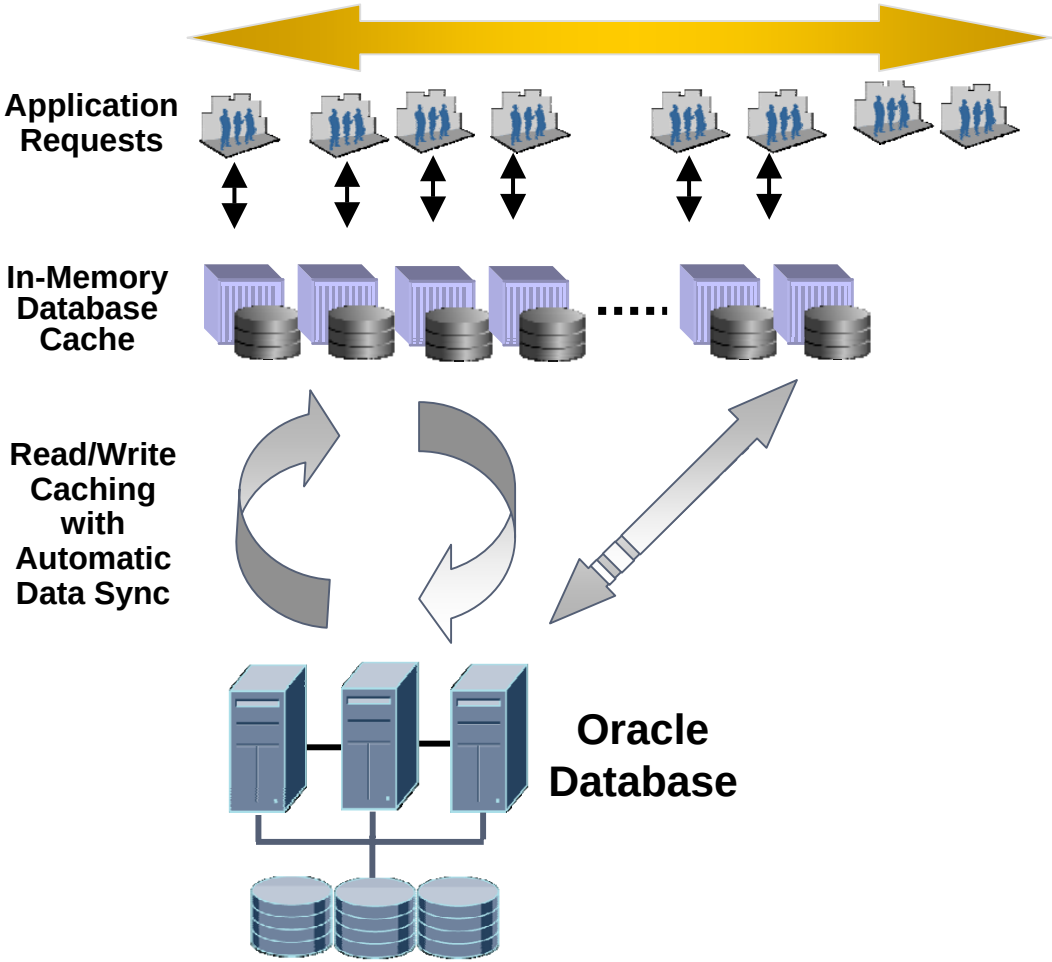
Simplest Queries can speedup:
- **50x in elapsed time**
- **20x in CPU time**

- Caches query results on client
- Primarily for caching small (10s or 100s of KB) read-intensive tables
  - Queries where network overhead dominates
  - e.g. lookup tables
- Cache is fully consistent
  - Coherence messages bundled into responses to DB calls ensure cache remain consistent
  - Like Cache Fusion extended out to clients

In-Memory Database Cache

Web Scale Highly Available

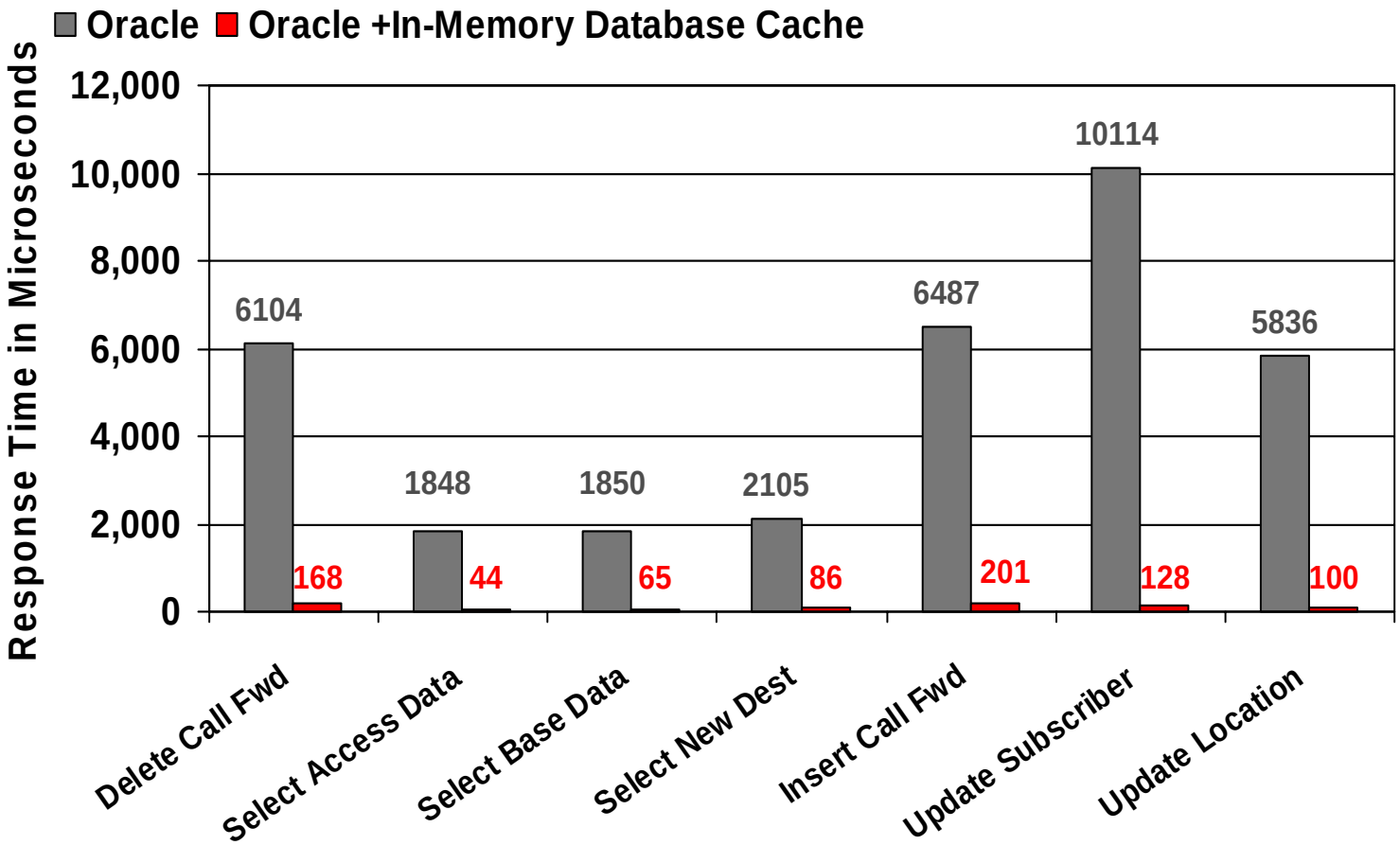
High volume real-time transaction and event processing



- Cache a subset of Oracle database tables in the application-tier
  - Improve response time
  - Unlimited read and write transactions
  - Automatic data synchronization
  - HA protection and no data loss
  - Scale with processing needs

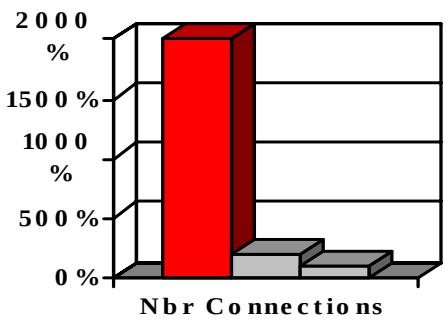
Blazing Fast Transaction Response Time

Leverage In-Memory Database Cache



Database Resident Connection Pool

Database Resident Connection Pool

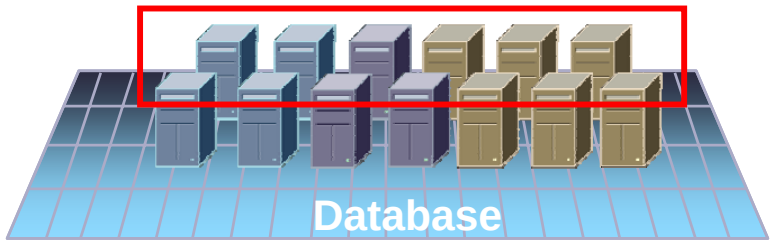


Up to 20 x connections

PHP

- Allows multiple Oracle clients to share a server-side pool of sessions
  - USERIDs must match
- Clients can connect and disconnect without the cost of creating a new server session and process
  - Similar to Web Server connection pooling model
  - Big scalability improvement for PHP applications
- Other networking improvements:
  - Network Fast Path for simple SQL performance
  - High throughput support for Bulk Data Transfers

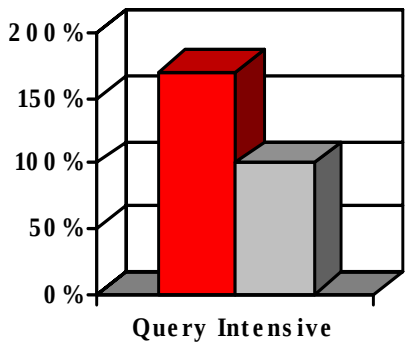
Adaptive Cache Fusion Protocols



- Next generation of cache fusion protocols.
  - Highly optimized for common operations
- Automatically selects best of several protocols by segment (table, index, partition) or query

Adaptive Cache Fusion Protocols

RAC Performance Enhancements

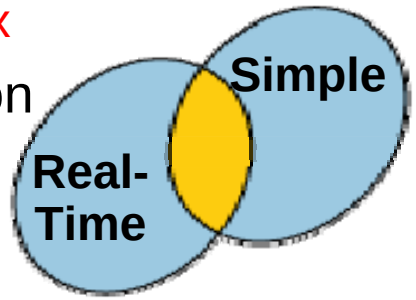


Up to 70% Faster

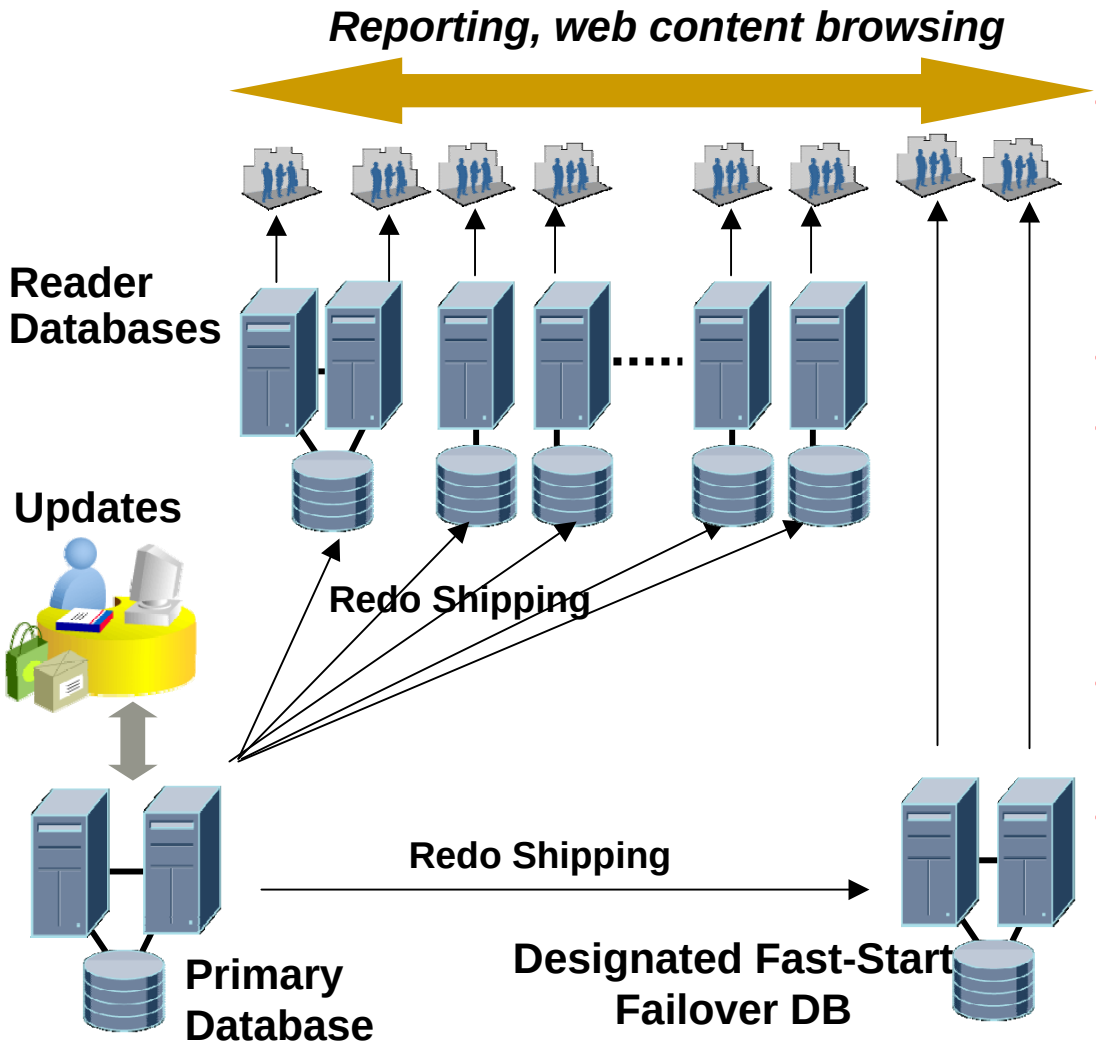
- Locality Optimized Protocol (10.2)
  - For segments primarily accessed by one instance
  - Eliminates messaging by that instance
- Read Optimized Protocol
  - For highly read-intensive segments
  - Eliminates messaging for readers
- Update Optimized Protocol
  - Apply update to block in parallel to readers releasing the block
- Table Scan Optimized Protocol
  - Large sequential table scans bypass caching which also bypasses messaging

Real-Time Reporting Database

- Users want to performance protect their production DBs
  - Reporting database offloads high risk reporting & backup from OLTP
- Current approaches -
  - Physical Copy Reporting DB (e.g. split mirror)
    - Solution is simple but data is stale (day old)
  - Logical Replica Reporting DB (e.g. replication)
    - Replication provides real-time updates but is complex
- **Active Data Guard** enables a unique real-time solution
  - Reporting using physical standby technology
  - Real-time, simple, and fast – also provides DR



Web Scale Highly Available Reader Farm

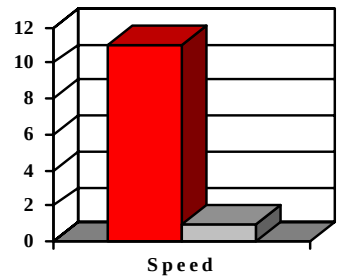


- Reader farm implemented using Data Guard Standbys
  - Scale-out read queries
  - Isolate faults to each DB
- 10g, use logical standbys
- 11g, can use Active Data Guard
  - High performance
  - Supports all types & DDL
- Automatic, zero loss failover
  - Readers follow automatically
- RAC can scale-out updater, or centralize storage of readers

“Native” PL/SQL and JAVA Compilation

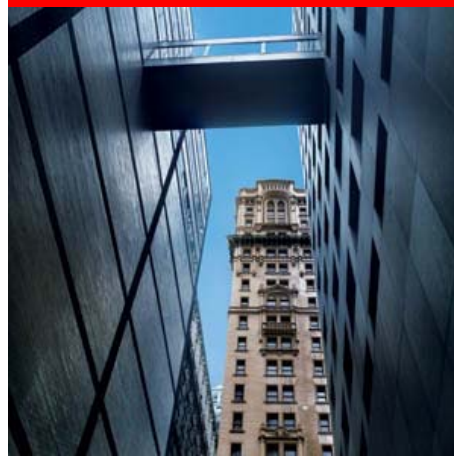
- Oracle compiles down to native instruction set
- 100+% faster for pure PL/SQL or Java code
  - 10% – 30% faster for typical transactions with SQL
- PL/SQL
  - Just one parameter - On / Off
    - No need for C compiler
    - No file system DLLs
- Java
  - Just one parameter – On / Off
  - JIT “on the fly” compilation
    - Transparent to user (asynchronous, in background)
    - Code stored to avoid recompilations

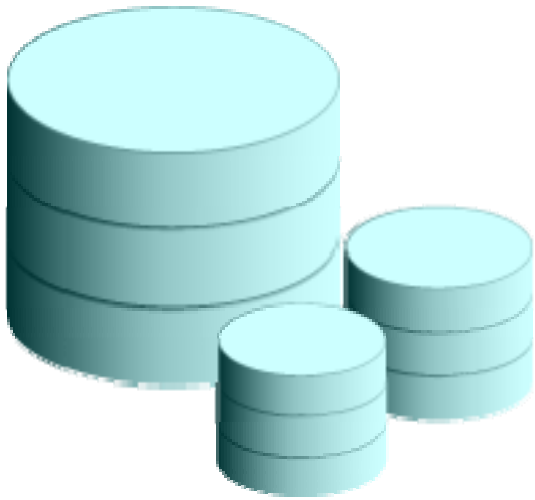
Java Just-In-Time Compiler



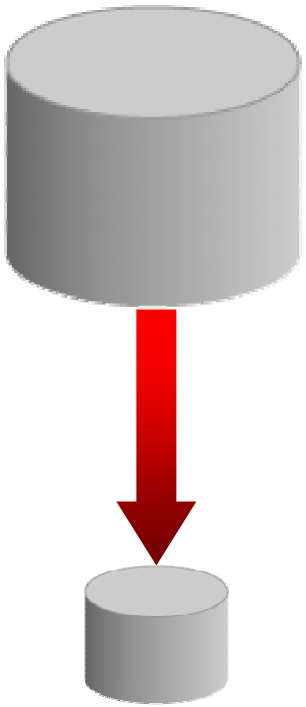
Up to 11 x Faster

Scalable Storage





Compression for Mainstream

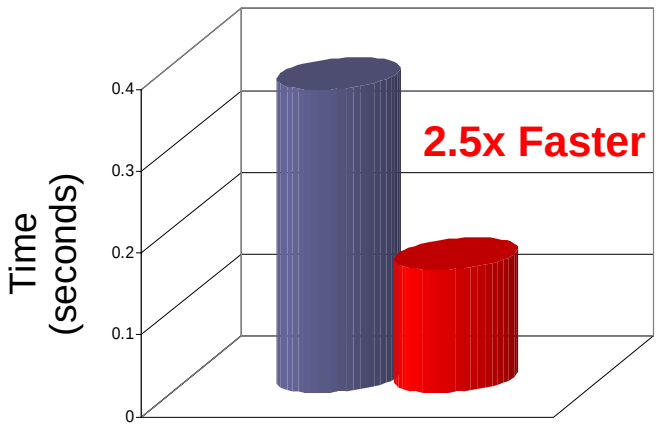


2x to 4x

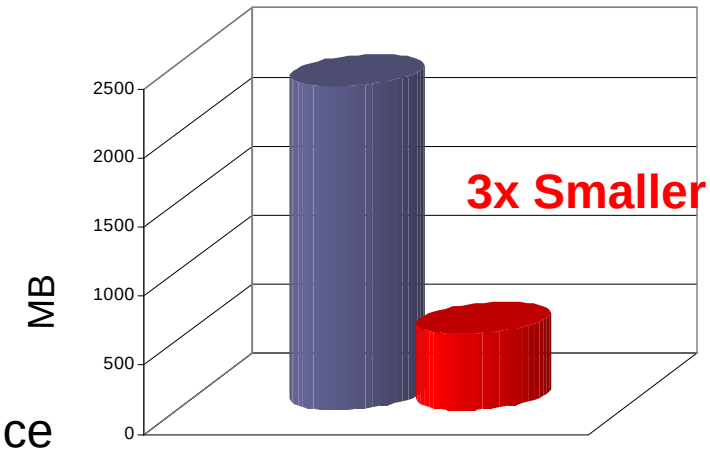
- Oracle 9*i* compresses table data only during bulk load
  - Data warehousing, ILM
- Table compression is now viable for all applications
  - Including OLTP Applications
- Typical compression ratio of 2x to 4x
- Database directly reads compressed data bypassing decompression overhead
- Strategy - Compress 10 largest tables in a database
  - Reduce table data by half, increase CPU usage 5%
- Savings cascades into test, dev, standby, mirrors, archiving, backup, etc.

Real World Compression Results - ERP Database 10 Largest Tables

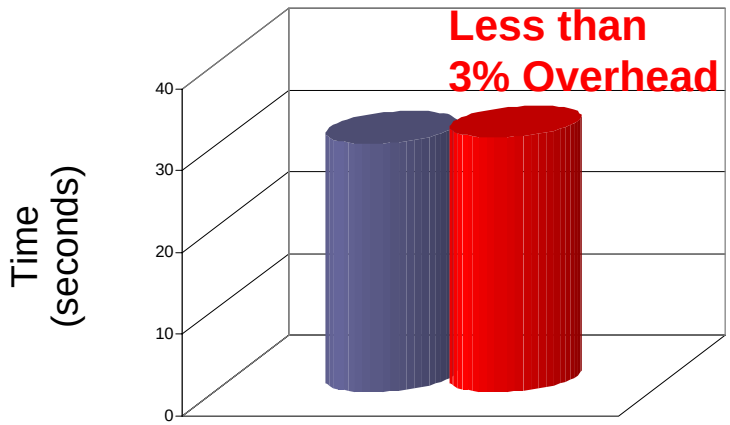
Table Scan Performance



Storage Utilization



DML Performance



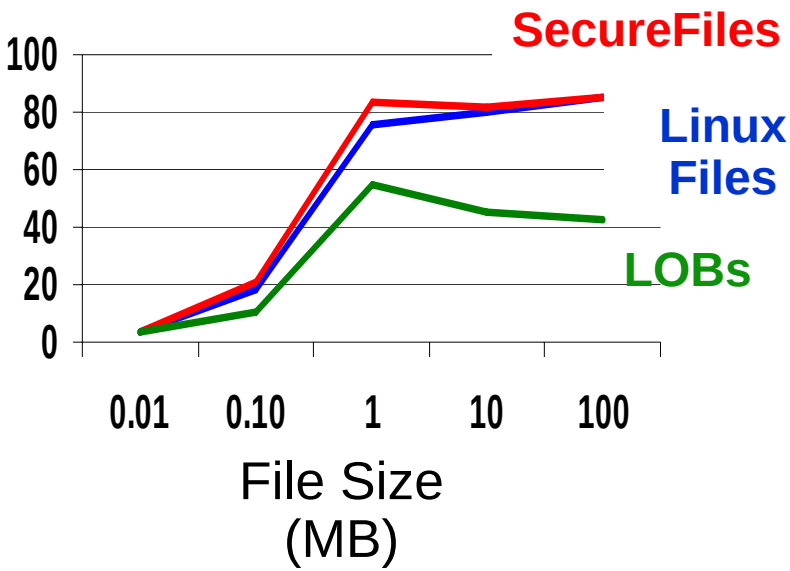
Oracle SecureFiles

Consolidated Secure Management of Data

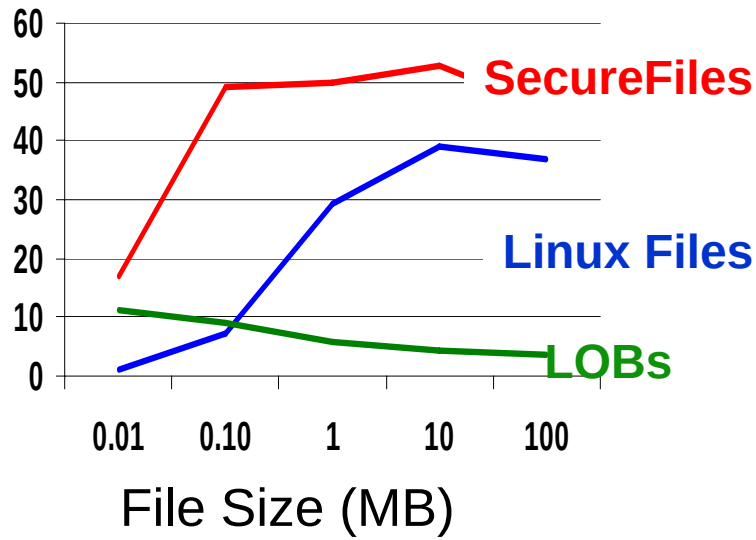
- SecureFiles is designed to break the performance barrier that has kept file data out of databases
- Similar to LOBs but much faster, and more capable
  - Transparent encryption, compression, deduplication, etc.
  - Preserves the security, reliability, and scalability of database
  - Superset of LOB interfaces allows easy migration from LOBs
- Enables consolidation of file data with associated relational data
  - Single security model
  - Single view of data
  - Single management of data

SecureFiles vs. Linux File System

File Reads
(MB/second)



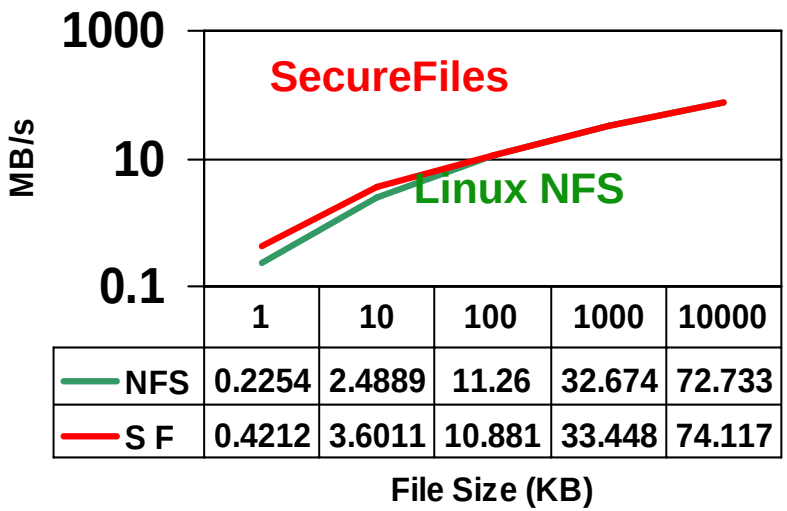
File Writes
(MB/second)



- Performance compared to Linux FS
  - Tests run using both SecureFiles and ext3 in metadata journaling only, no network

SecureFiles vs. NFS

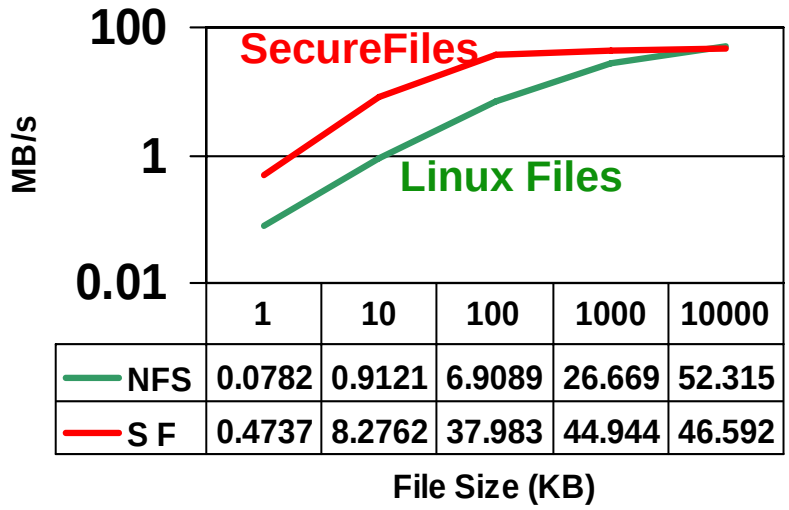
Read Performance



	1	10	100	1000	10000
NFS	0.2254	2.4889	11.26	32.674	72.733
S F	0.4212	3.6011	10.881	33.448	74.117

File Size (KB)

Insert Performance



	1	10	100	1000	10000
NFS	0.0782	0.9121	6.9089	26.669	52.315
S F	0.4737	8.2762	37.983	44.944	46.592

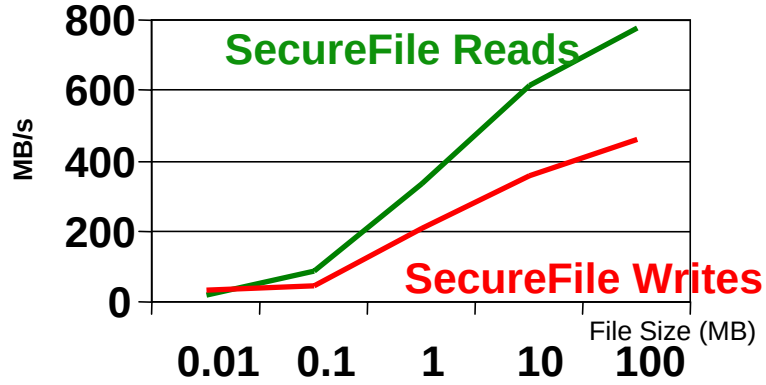
File Size (KB)

- Performance compared to Linux NFS
  - Up to 2x faster for Queries, 6x for Inserts
  - Tests run using both SecureFiles and NFS/ext3 in metadata journaling only (default for NFS)

SF: Breaking The Performance Barrier

- **High Performance**
  - **38TB/day injest**
    = 5x YouTube
- **Unlimited Scalability**
  - **RAC for Server**
  - **ASM for Storage**
- **SecureFiles is free**

High Performance Experiment



- 776 MB/s for File Read
  - **67TB/day of data serve**
- 462 MB/s for File Writes
  - **38 TB/day of data ingest**
- 8 sessions, 4 node RAC, x2 Xeon, 6GB RAM, 3 EMCCX700

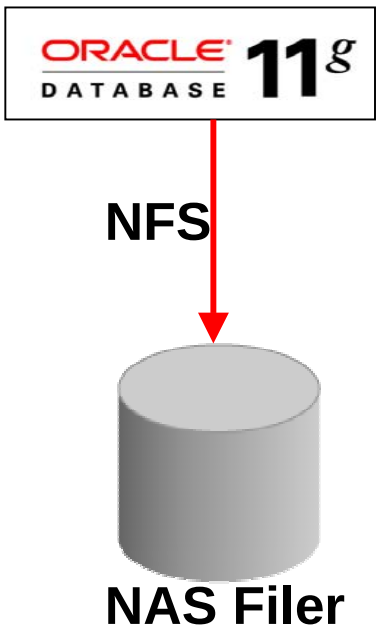
Designed to Scale

Virtual Columns & Partitioning

```
Create table t1 (
    c1 number, c2 number,
    c2 as (c1+c2) virtual)
```

- ANSI syntax
- Look just like regular columns from SQL perspective
- Support for partitioning, indexes, constraints, statistics, histograms
- Used by expression evaluation when applicable

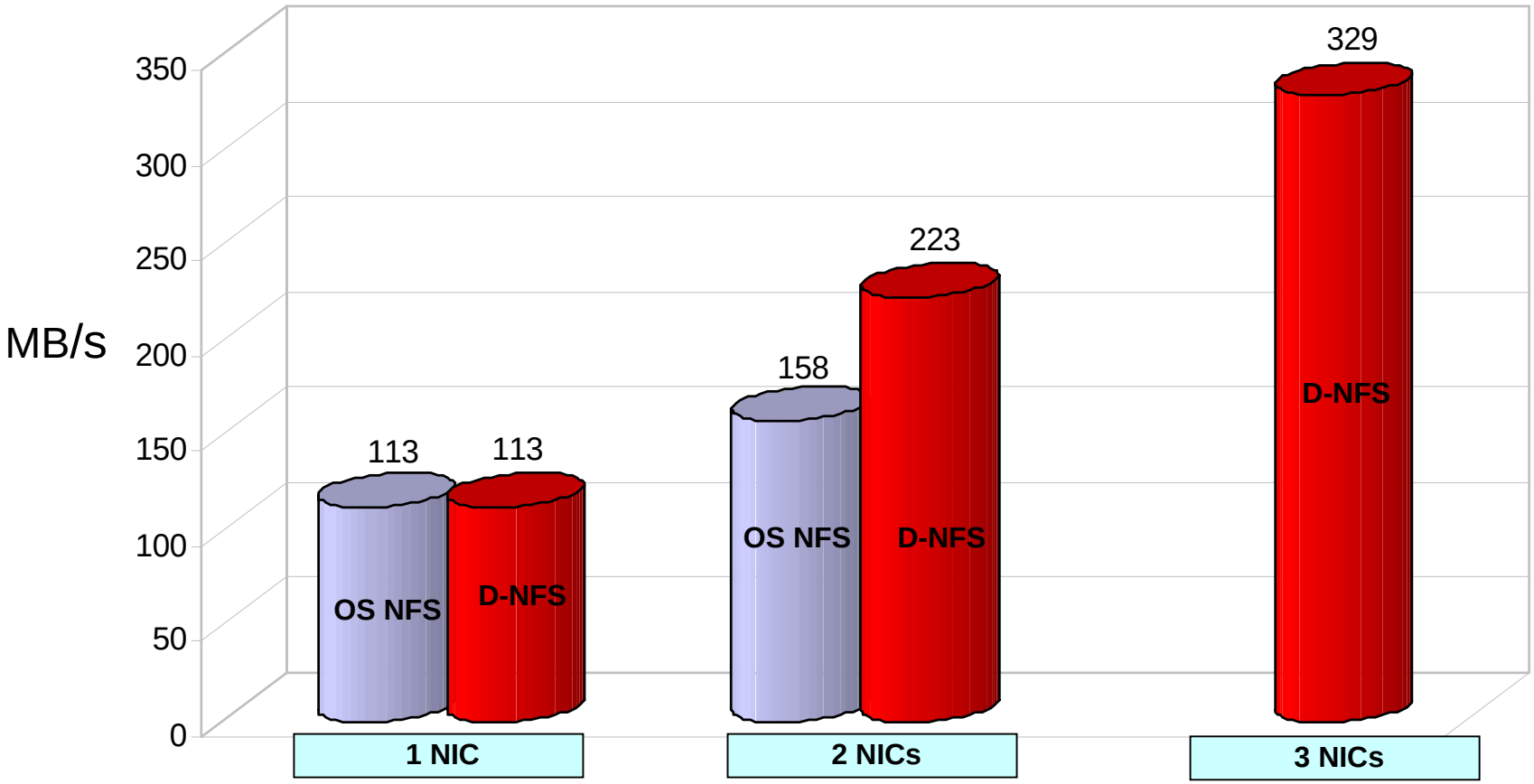
Direct NFS Client (D-NFS)

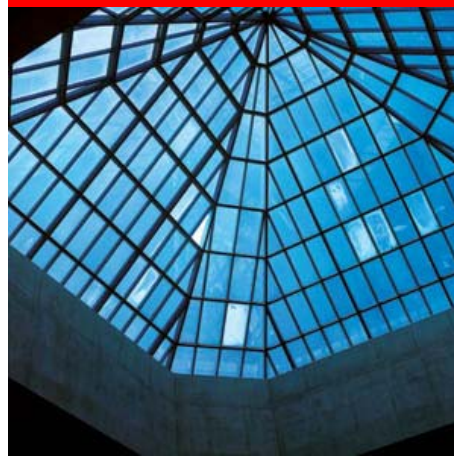


- Oracle directly talks NFS V3 protocol to NAS Filers for database files
  - Bypasses OS NFS client implementation
  - Runs on all Operating Systems including Windows
  - Simple - preconfigured and optimized for Oracle
  - Relief from OS NFS client complexities, bugs
  - Automatic link aggregation

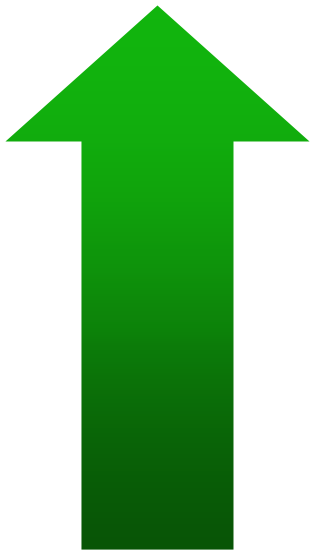
Direct NFS Client (D-NFS)

Throughput: OS-NFS vs. DNFS



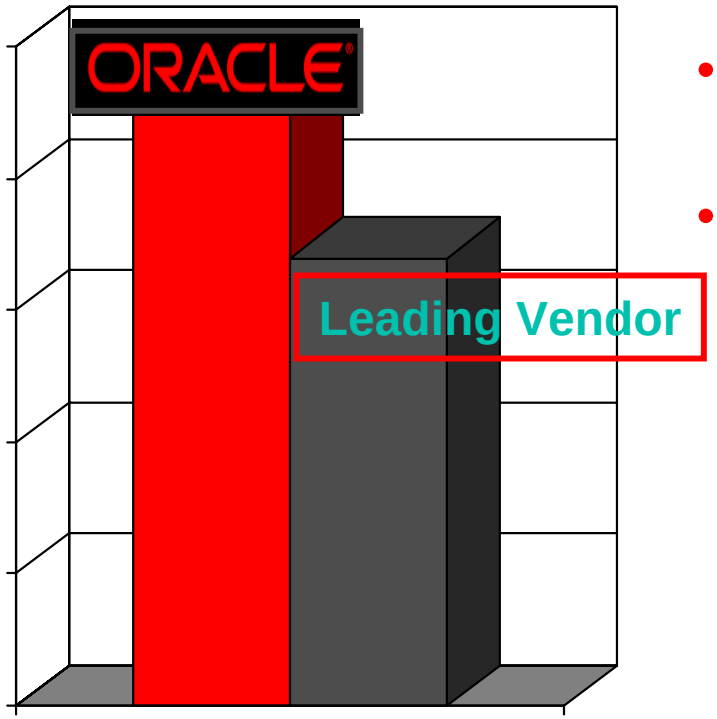


Scalable Availability



Scalability aspects of HA technologies

World's Fastest Database Backup

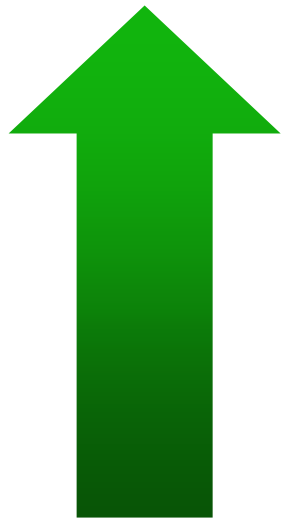


Speed of Database Backup

- Oracle Secure Backup is fast
  - 15% to 30% faster than competition
- Only backup that is tightly integrated with Database kernel
  - Less overhead with direct calls into DB engine
  - Eliminates backup of unused space
    - DB 10.2
  - Eliminates backup of committed UNDO
    - DB 11g

Faster Backup & Recovery

- Intra-file parallelism for backup and restore of Bigfiles
- Fast Incremental Backup on Physical Standby Database
  - Tracks changed blocks on standby database
- New parallel media recovery doubles redo apply performance
- Fast Streaming Data Guard Redo Transport
  - Transport does not wait for acknowledgement
- Data Guard SQL Apply and Streams performance improved up to 50%

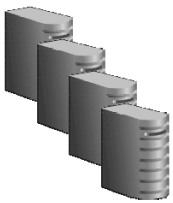


Scalable Management

Management at Scale

- Manage Highly Complex Integrated Applications
  - Partitioning Advisor
  - Fully Automatic SQL tuning
  - ADDM for RAC
  - Streams Performance Advisor
  - Capture/Replay for Database Workloads
  - Capture/Replay of high load SQL
  - SQL Plan Management – controlled plan evolution
  - Automatic Performance Baselines and Metric Thresholds

Oracle is Ready to Scale Today & Tomorrow



- **Scalable Execution**
  - Query Result Caches
  - Reader Farms
  - Enhanced Native Compilation
  - Connection Pooling
  - Optimized Cache Fusion Protocols

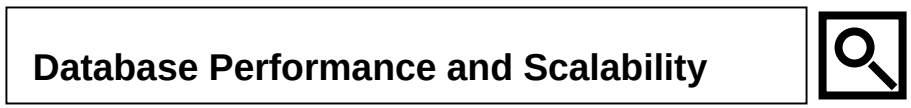
- **Scalable Availability**
  - Oracle Secure Backup
  - Faster Backup & Restore, Data Guard
  - Readable Physical Standby

- **Scalable Storage**
  - Table Compression
  - SecureFiles
  - Partitioning by Interval, Virtual Column, Ref, Composite

- **Scalable Management**
  - Workload Capture & Replay
  - Performance Advisors
  - Plan Management

For More Information

http://search.oracle.com



or

http://www.oracle.com/


ORACLE®