



# Oracle Engineered Systems at Thomson Reuters

Engineered Systems – the Foundation of Efficiency

Aaron Pust

April 16th

ORACLE®



THOMSON REUTERS

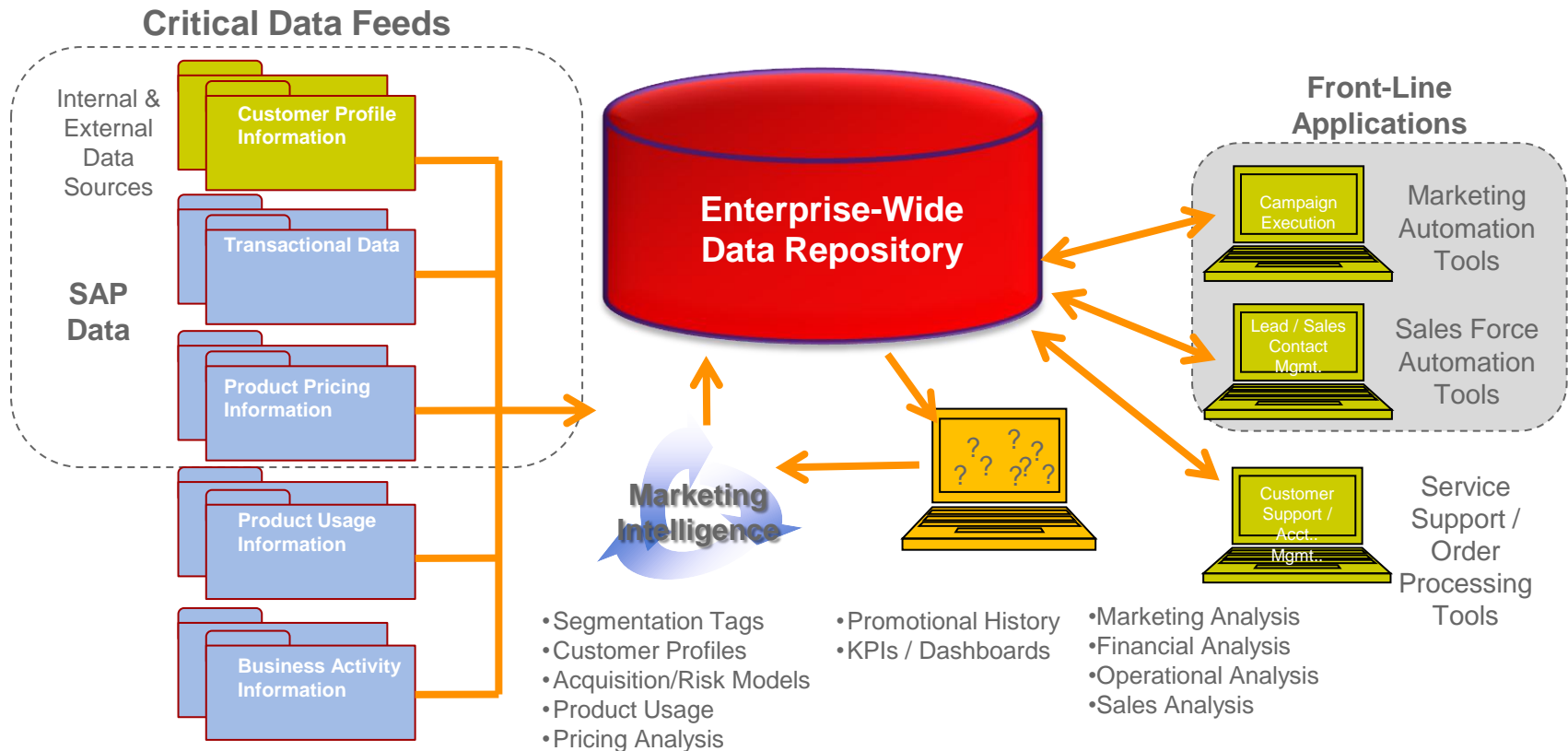
# Engineered Systems Strategy



- Thomson Reuters – use cases in an information company
  - Enterprise Data Warehouse – Customer, product, and usage data analysis
  - Risk & Fraud Warehouses - Product serving up large amounts of information
- Large scale Data Warehousing workloads
  - Large growth of data warehouses and the need to load, analyze, and report on huge data sets quickly
  - Stretched traditional approaches to data warehousing to solve the throughput bottlenecks
  - High costs to deploy and optimize
- Oracle Solution Proposition - Exadata
  - Delivery strategy as a Turnkey solution of balanced configuration with both compute and storage
  - Unique software solution in the storage tier
- Evolution of the Oracle Solution at Thomson Reuters
  - Early Adopter

# Enterprise Data Warehouse

Enterprise repository of information on our customer and their interactions with our products. This platform supports the various analytics used by marketing and other business operations



# Enterprise Data Warehouse – Next Generation

---

- Next Generation of the Enterprise Data Warehouse for Business Systems is Needed
  - Mixed workload, and concurrency requirements severely impacting response times for all users.
  - Current environment unable to support response time requirements of the sales support projects.
  - 3Cs – Concurrency, Consistency, and Cost
    - **Concurrency** – Increase capacity in data size and number of users
    - **Consistency** – Increased and predictable performance
    - **Cost** – Cost effective solution that can scale to meet future needs
- Trials with IBM, Oracle, and Greenplum
  - Replace a traditional warehouse solution using IBM p-series servers and Hitachi SAN storage (P570 12 cores 192GB)

# Evolution of the Oracle solution



**Exadata V1**  
Introduction of  
Oracle  
Database  
Machine Oracle  
11.1



**Enterprise Data  
Warehouse**  
Initial Rollout on  
Oracle Exadata  
V1



**Exadata  
Benchmarks**

*Evaluation for MIS Data  
Warehouse platform  
(Oracle Labs)*

**ORACLE®**

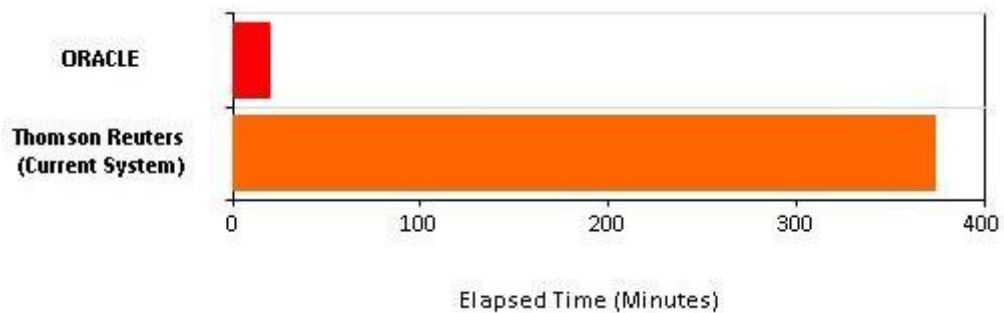


**THOMSON REUTERS**

# Enterprise Data Warehouse Evaluation of Alternatives

- Test Setup
  - 43 queries taken from the production system
  - Mix of long, medium, short running Business Objects and Siebel Analytics dashboard queries.
  - Custom developed test driver program used to simulate activity in current environment.
  - 3.2 Terabytes of usage and revenue data, nearly half total size (7.5 billion rows).
- Test Execution
  - Baseline Query Execution
  - User Simulation Stress Test
  - Mixed Workload Stress Test

**Baseline Test = All 43 benchmark queries run single file**



Example Long Queries	TR	ORACLE
	Minutes	Minutes
BO_Q6	15.98	0.38
BO_Q17	28.31	1.01
BO_Q2	29.29	0.60
BO_Q3	37.20	0.87
BO_Q15	77.87	1.72
BO_Q21	90.16	1.94

❖ Oracle more than 19x better than current system.

# Enterprise Data Warehouse Evaluation of Alternatives

Significant improvement in **Concurrency**, **Consistent** performance in mixed workloads, **Cost effective** choice in an all-in-one model.



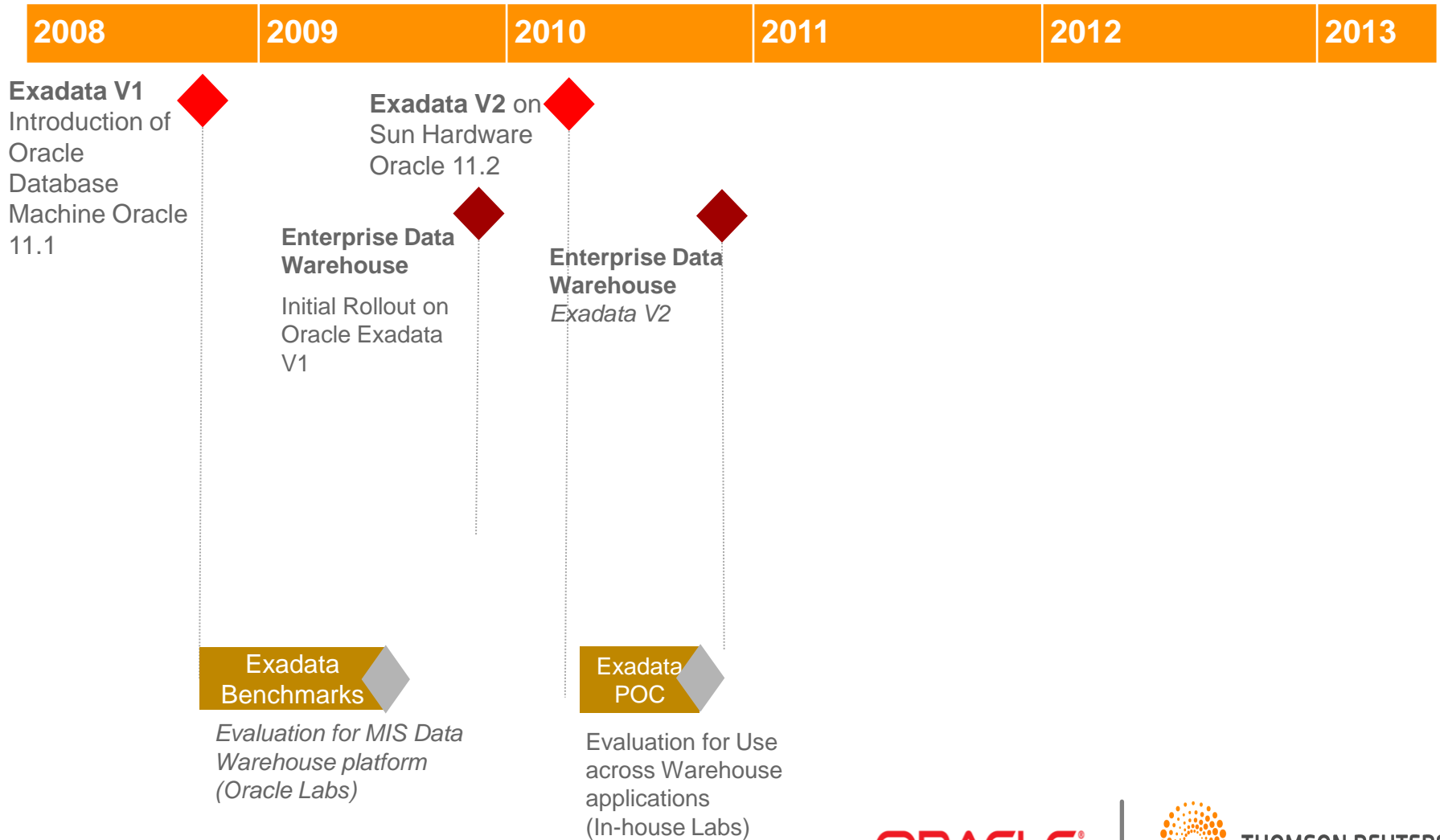
- Long Queries
  - TR system unable to complete a long query during these 3 hour tests.
- Medium Queries
  - Over 5x advantage for Oracle.
- Business Objects Short Queries
  - 4x advantage for Oracle over current system.
- Siebel Analytic Short Queries
  - Somewhat anomalous data skew in this category, however, still extremely wide gap for Oracle.
- ETL Operations
  - 5x advantage for Oracle.

ORACLE®



THOMSON REUTERS

# Evolution of the Oracle solution





# The Compression Advantage

## Exadata V1 – OLTP Compression

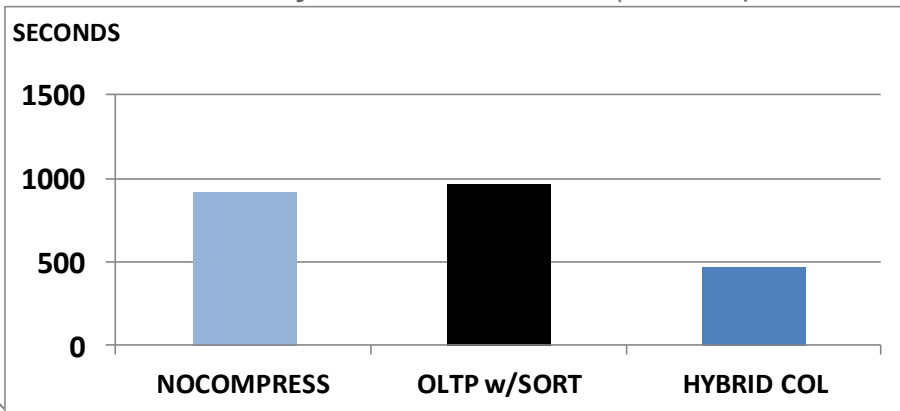
- 7 Large Table sample
- 2.5:1 – No Compression
- 5:1 – OLTP

7 large Fact Tables	Initial Size	V1 - OLTP
RAW Data Size	4.48 TB	2.96 TB
Compressed Data Size	<u>1.75 TB</u> (2.5 : 1)	<u>584 GB</u> (5 : 1)
Index Space*	2.3 TB	148 GB
Total Used Space	4 TB	<b>732 GB</b>

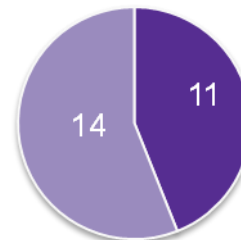
## Exadata V2 – Hybrid Columnar Compression

- 10:1 or Higher with proper sorting

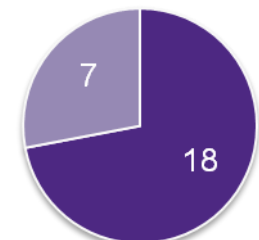
EXADATA v2 Query Performance Test ( 25 test queries run single file with no indexes )



OLTP vs. NOCOMPRESS



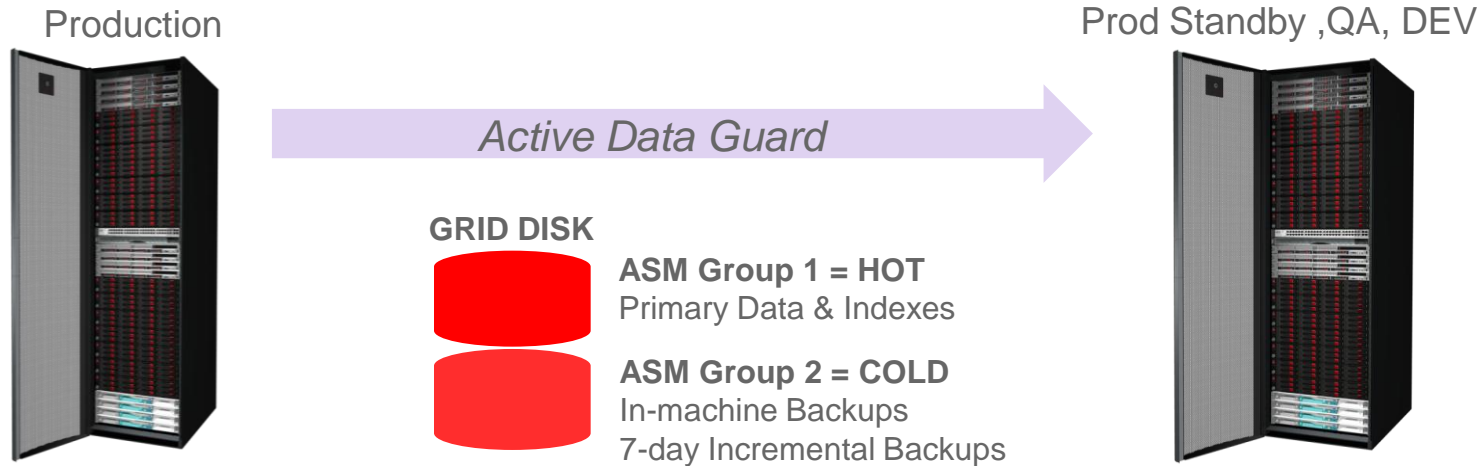
Hybrid Columnar vs. NOCOMPRESS



■ - Queries faster with compression

# Backup Strategies

Initial Strategy kept data Backups on Exadata Storage



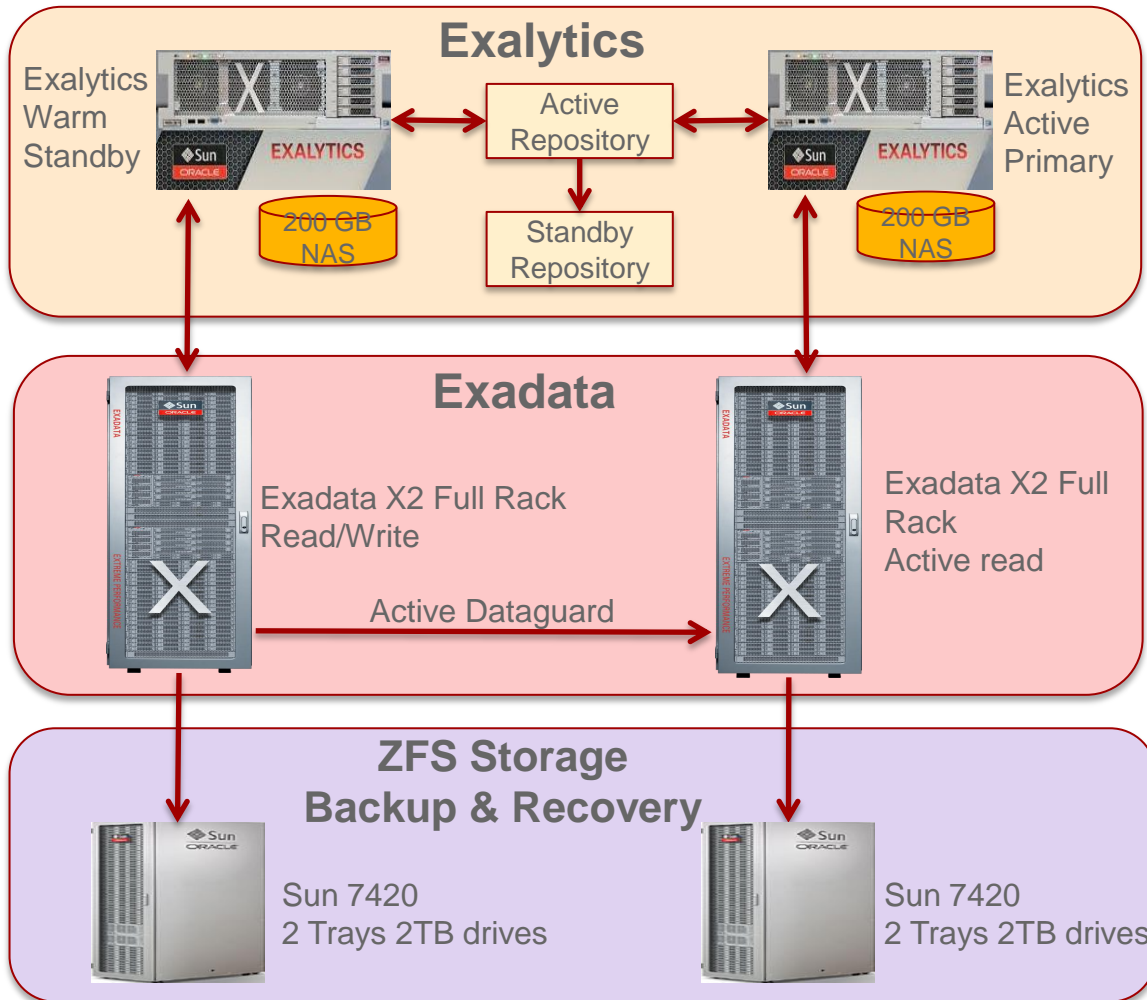
Introduction of Sun ZFS Storage Appliance to enable off machine backups using RMAN

- Enabled expansion on Exadata for Production Workload
- High Speed Interconnect with Dedicated backup appliance
  - 9.8 TB backups in 1 hr 15 min
    - Incremental is 15 min



Sun 7420  
2 Trays 2TB drives

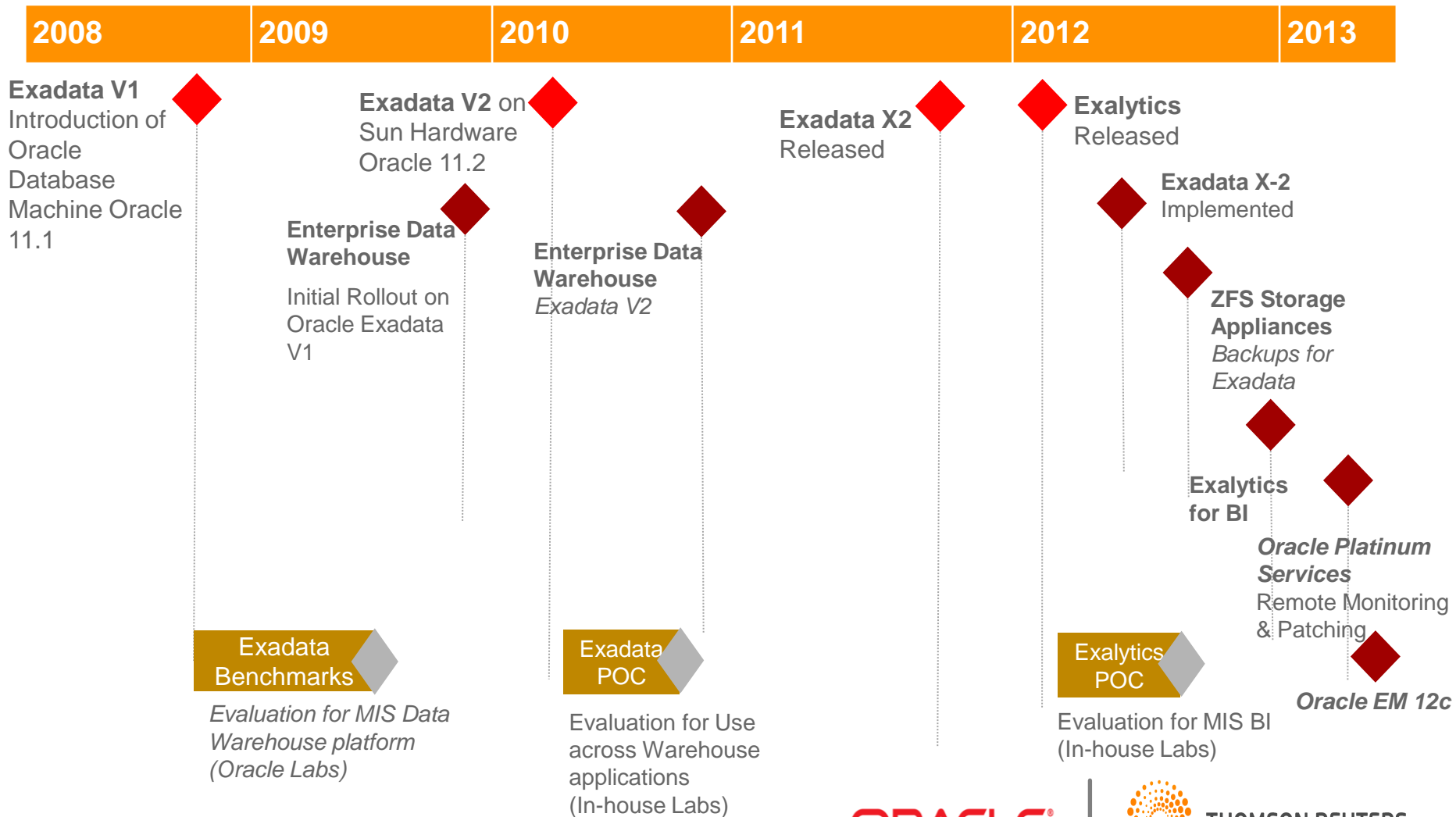
# Enterprise Data Warehouse Implementation Architecture



## EDW STATS

Raw Data Size	20 TB
Compressed Data Size	5 TB (4:1) Max(12:1)
Index Space	700GB
ETL Batches	900
Users	~8000
Concurrent Sessions	~300
Tables	~3,500
Indexes	2,368
Largest Table	200 GB 20 Billion Rows
Data Change Rate	500 GB/day

# Evolution of the Oracle solution

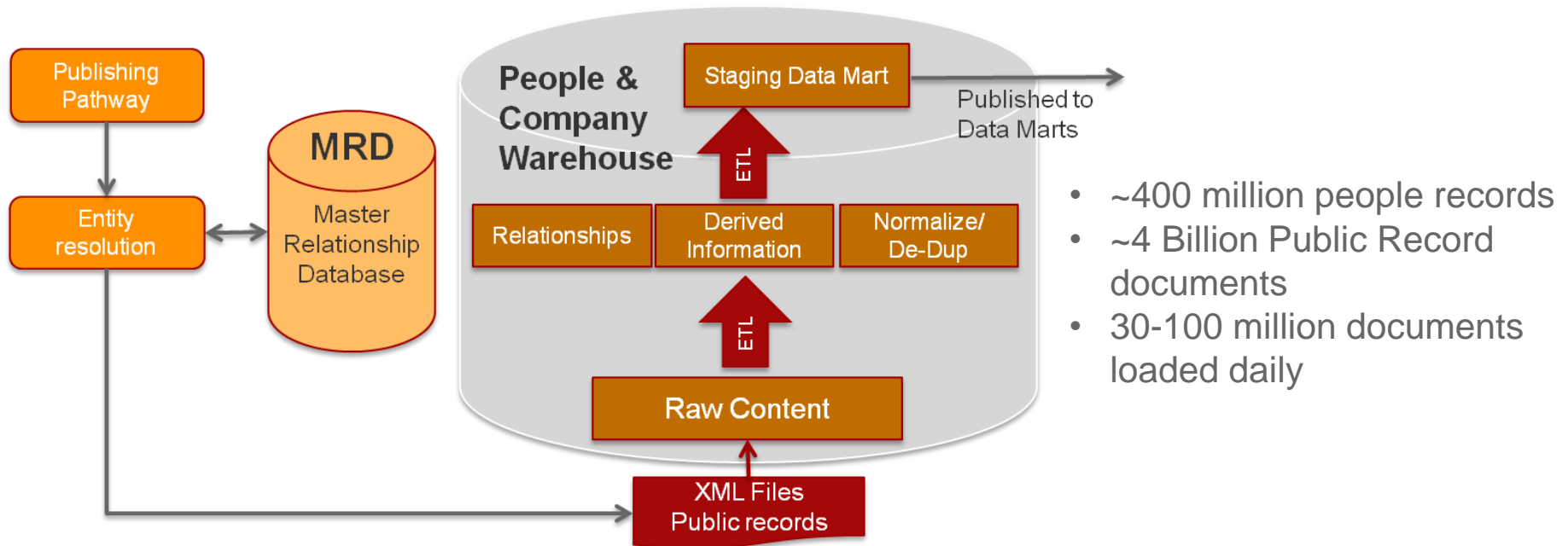


# Risk & Fraud – Data Warehouses

Need for a new data warehouse for public information requiring significant performance and scalability

Large amounts of data processed for relationships requiring high performance

- Large data sets to identify connections and recognize relationships between people
- ETL processes to support the creation of data marts for various products in our Risk & Fraud business



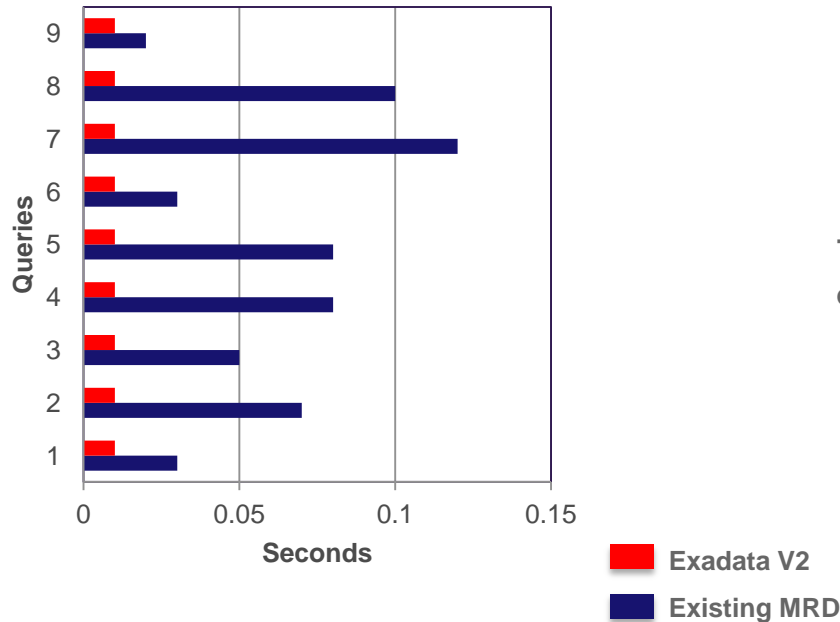
# Risk & Fraud Warehouses – Evaluating Exadata

---

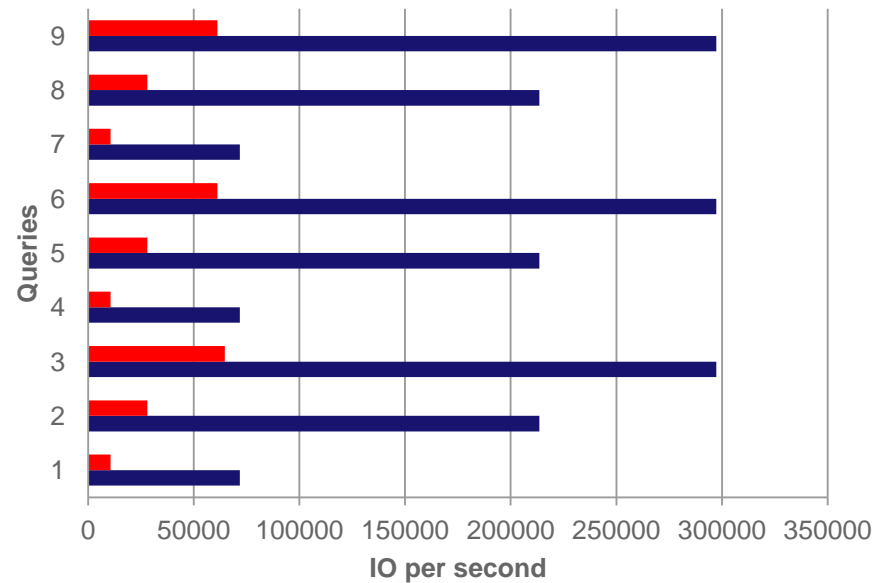
- Testing focused to determine if Exadata could support the growing requirements of the Master Record Database (MRD).
  - Ability to cost effectively double the current size
    - Existing system was memory and CPU constrained
      - IBM P-series P570 16 cores/CPU and 768GB memory
  - Support 3x the number of concurrent Connections
    - Needs to support more entity resolution processes
  - Able to balance multiple data warehouse activities without impeding performance
    - Must be able to keep up with the loads and monthly updates
  - Future plans for expansion will significantly increase the number of documents to be processed
    - Acquisition of new content
    - More Historical records for individuals
    - Addition of deceased people

# Risk & Fraud Warehouses – Evaluating Exadata

## Elapsed Time per Execution



## Logical IO per Second



### Queries

Test 1 (160 threads) = 1,4,7  
 Test 2 (320 threads) = 2,5,8  
 Test 3 (480 threads) = 3,6,9

- Exadata was 2-12x faster than production MRD
- Performance degradation on Exadata was minimal as we scaled up to 3x concurrent connections
- Logical IO per second with Exadata is ~16% of the amount on MRD

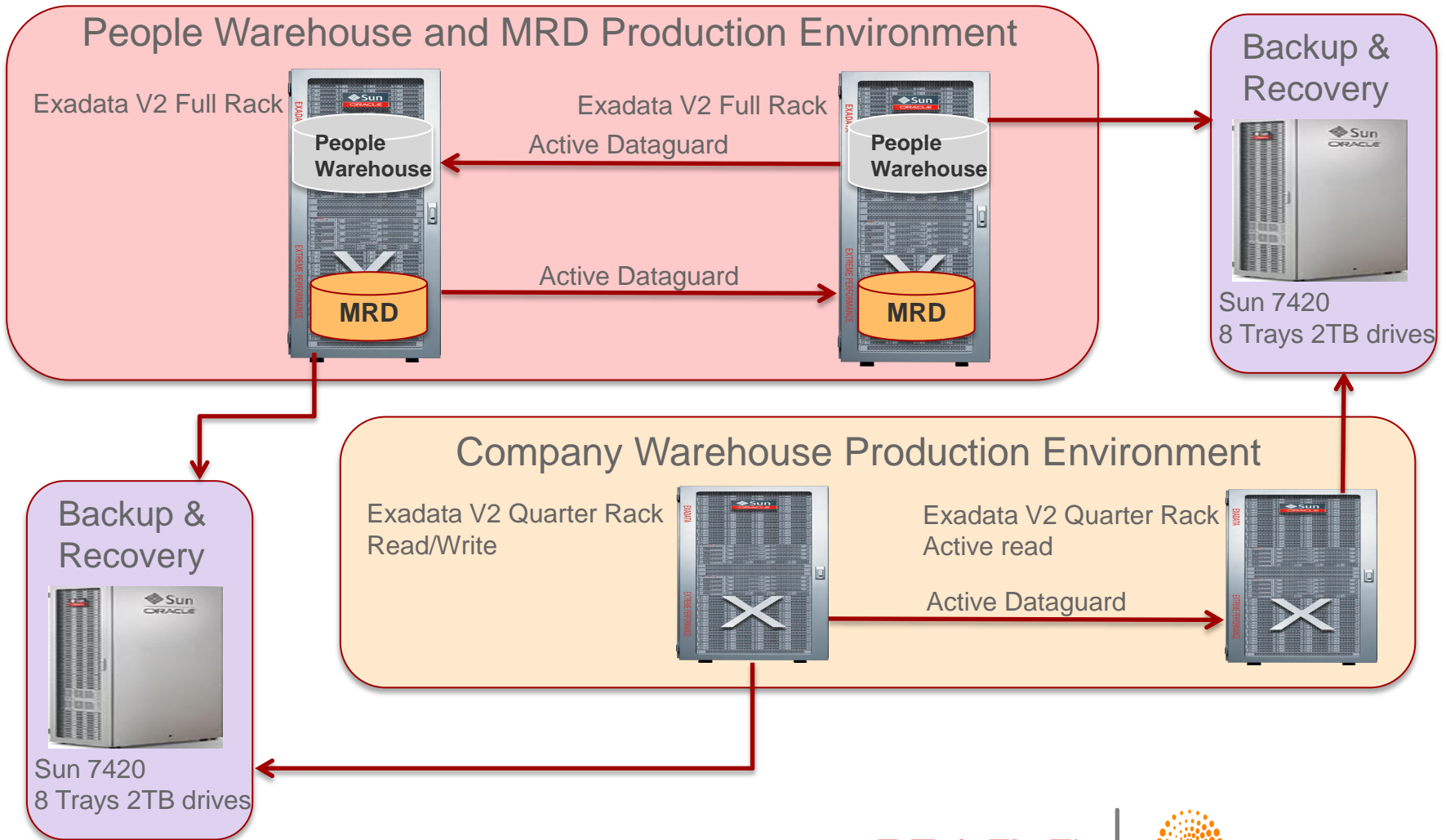
# Risk & Fraud Warehouses – Evaluating Exadata

---

- Exadata continually outperformed existing MRD in our tests
  - 2-12x faster “out of the box” with minimal tuning on Exadata
  - Redeploying on Exadata will immediately increase processing capacity of the publishing pathway
  - Proved that increasing the size and connections to MRD on Exadata while simulating other ETL activities still was faster than current systems
  - Removing existing Materialized Views solution enabling faster updating
    - Indexes were able to fit into the cell flash cache and exceed performance goals
- Exadata workload management facilities allowed for a much more balanced system behavior
  - Allows for resources to be utilized for other ETL activities



# Risk & Fraud Warehouses Implementation Architecture



# Risk & Fraud Warehouse Statistics

## People Warehouse STATS

Data Size	19TB
Index Space	112 GB
ETL Batches	13
Concurrent Connections	104
Tables	4,188
Indexes	98

## Company Warehouse STATS

Data Size	4.5 TB
Index Space	608 GB
ETL Batches	300
Concurrent Connections	60
Tables	870
Indexes	199

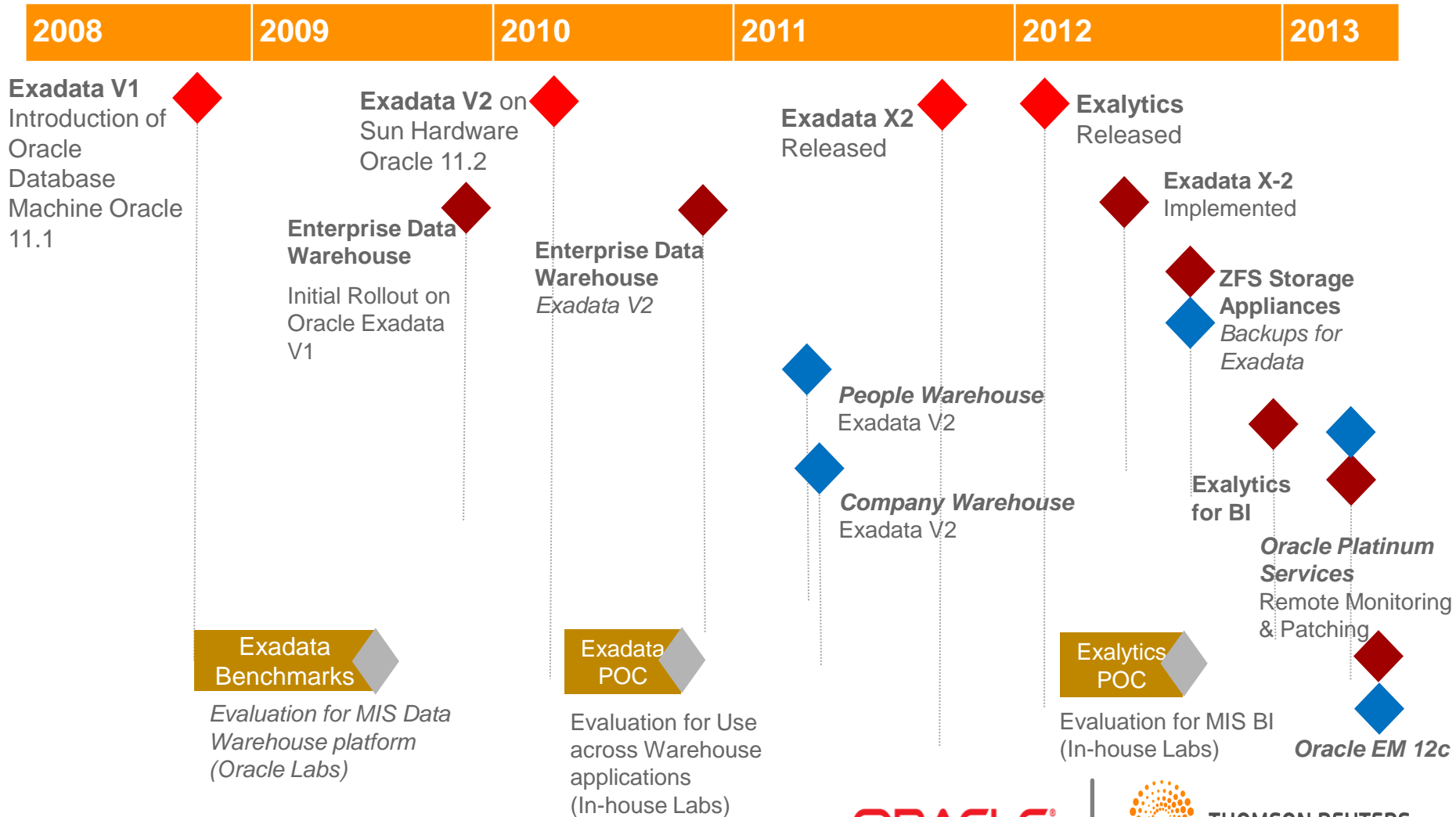
## MRD STATS

Data Size	1.3 TB
Index Space	3.4 TB
ETL Batches	12,000
Concurrent Connections	12,000
Tables	118
Indexes	297

## High Data Guard Traffic

- Average around 8-10TB of archive log generation per day for People Warehouse
  - Peak was 17TB
- Backup Timings RMAN to ZFS Storage Appliance
  - People Warehouse – 13 hours for Full Backup
  - Company Warehouse – 6 hours for Full Backup

# Evolution of the Oracle solution



**ORACLE®**



**THOMSON REUTERS**

# Conclusions

---

- Exadata Appliance Model
  - Secret Sauce – the storage layer
    - Smart Scans to reduce data sent back to the Database
    - Smart Flash Cache to increase scan bandwidth
    - I/O Resource Manager to prioritize I/O and enable predictable performance
    - Hybrid Columnar Compression
    - 40Gb InfiniBand networking
  - Still requires effort
    - Initial stabilization was a hurdle with many patches and configuration changes
    - Application changes can enable greater performance
      - Significantly decrease materialized views and indexes
      - Compression gains by changing load methodology
    - DBA resource efforts moved away from performance tuning to administration
    - Support Model is cross discipline
    - Patching dependencies across all tiers (Hardware, OS, DB, Storage)
  - Growth Increments are restricted in the model
  - Data Center Space and Cooling
    - Engineered Systems need to be close by each other

# Thank you

---

*Scan this QR-code to add me to your contacts*

Contact: [aaron.pust@thomsonreuters.com](mailto:aaron.pust@thomsonreuters.com)

