

ORACLE®

ORACLE®

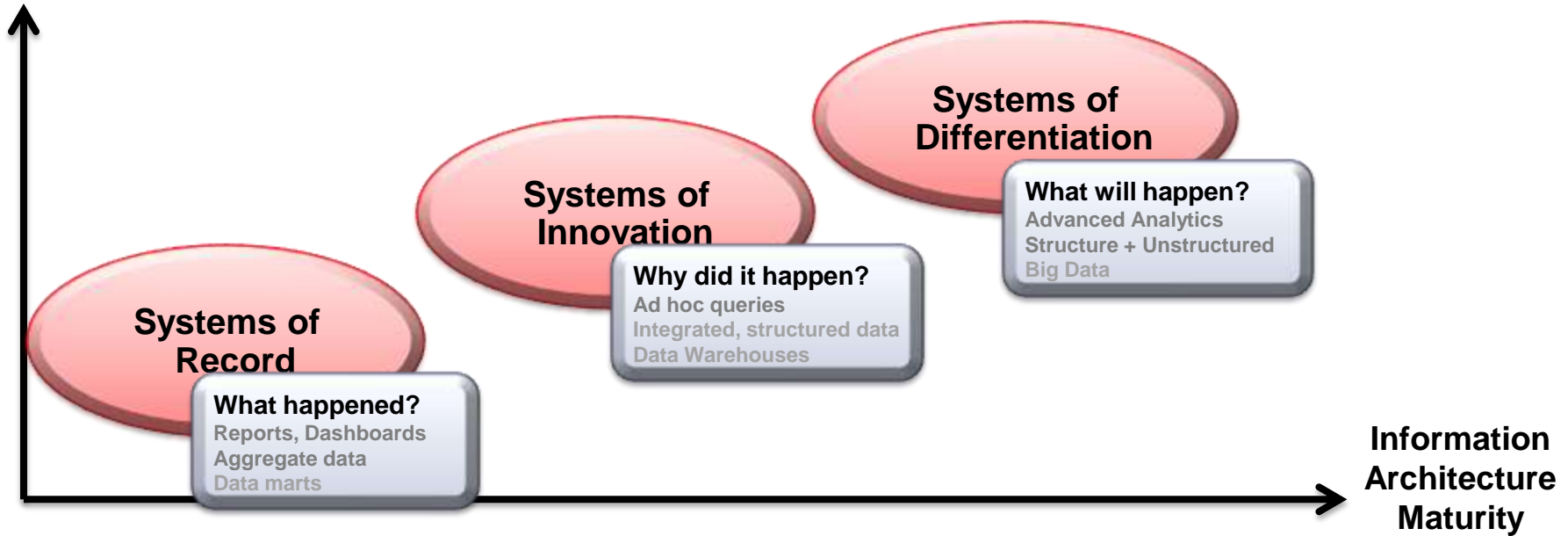
Data Warehousing and Big Data with Oracle Database 12c

George Lumpkin
Vice President, Product Management
Oracle Database Server Technologies

ORACLE®
DATABASE 12^c

Taxonomy of Information Architectures

Business
Value



Categorizations based up Gartner Pace-Layered Model: <http://www.gartner.com/technology/research/pace-layered-application-strategy/>

ORACLE

Optimized for systems of record



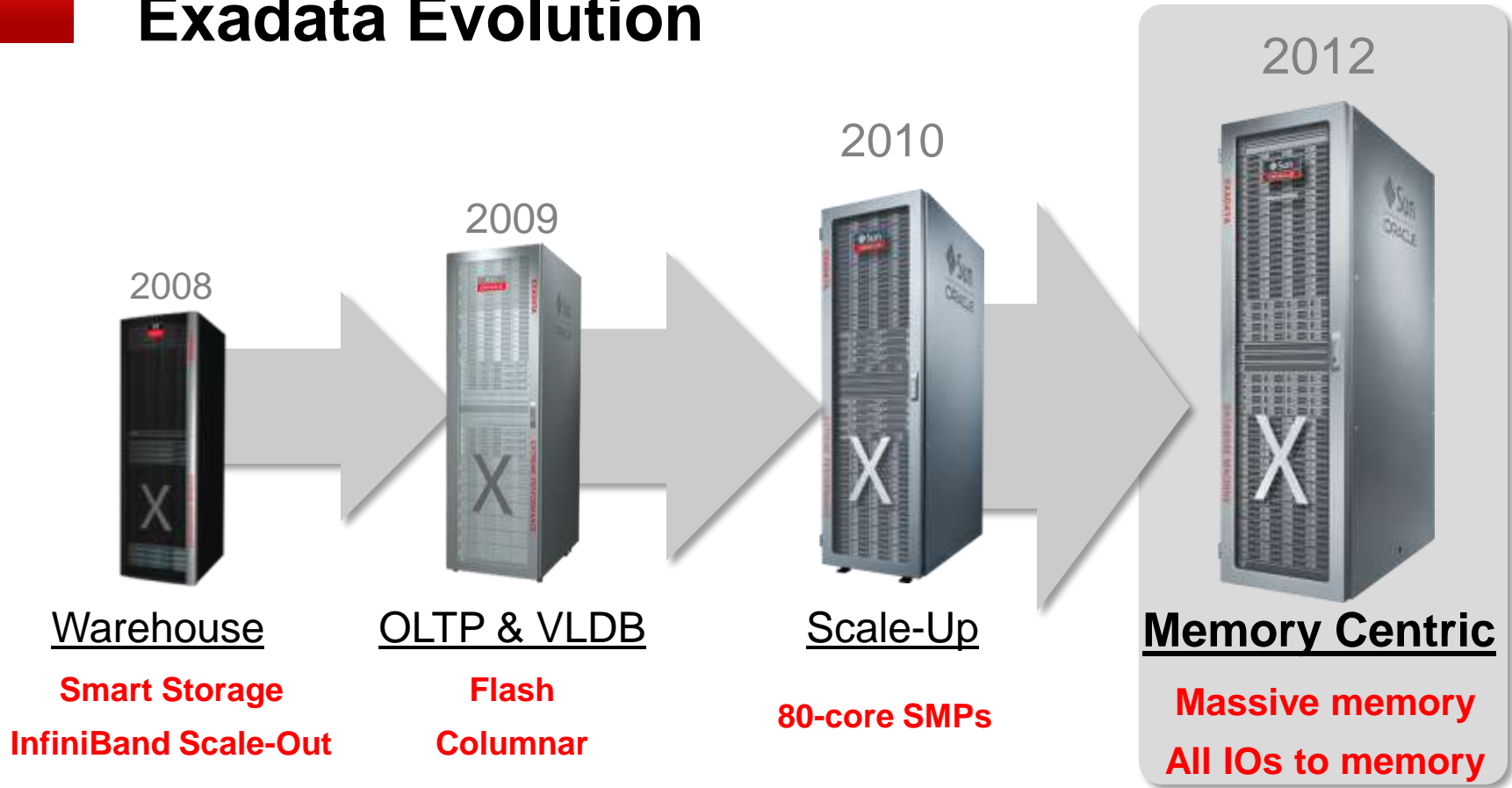
- Example applications: Data marts, Dashboards, BI-enabled applications
- Key Requirements:
 - Fast response times
 - Predefined reports – but potentially complex queries
 - Large numbers of concurrent users
- Key technical capabilities
 - Advanced indexing
 - Dimensional schema optimization
 - Aggregate management

Exadata for Systems of Innovation



- Optimized for large-scale ad hoc queries over large volumes of detail data
 - Partitioning
 - Parallel Execution
 - Exadata Hybrid Columnar Compression
 - Exadata Smart Scans
- Exadata scans data at much faster rates
 - Fast scans from disk, flash, and memory, all fully parallelized

Exadata Evolution



New Features for Data Warehousing

Performance and scalability

- Adaptive Query Optimization
 - Adaptive plans
 - Adaptive statistics
- New table partitioning technique: interval-reference
- Parallel Execution enhancements
 - Concurrent Union-All
 - Parallelization of correlated filters and expressions

“Always on” Data Warehouse

- New online DDL operations: partition move and others
- Asynchronous partitioned global index maintenance
- Invisible columns
- Out-of-place refresh and synchronous refresh for materialized views

Data scientists

- SQL Pattern Matching
- Advanced Analytics
 - Enhanced “Data Miner GUI
- New in-database predictive algorithms
- Further embedding of R capabilities into database

Adaptive Query Optimization

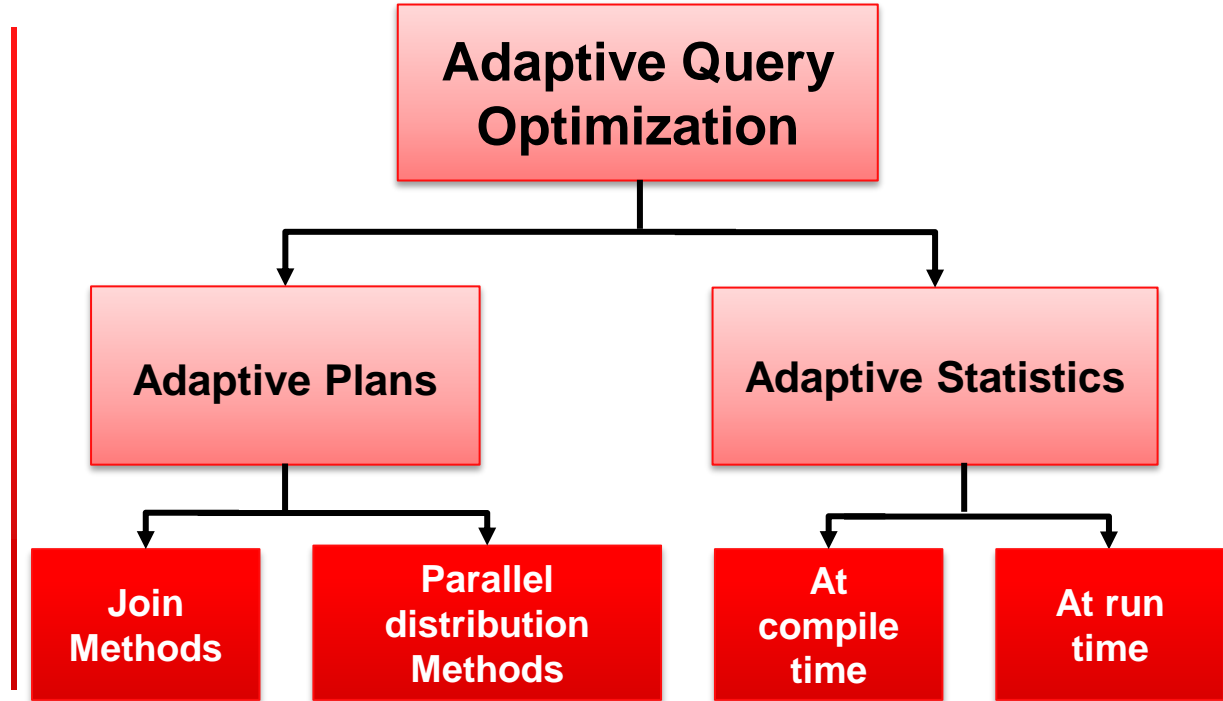
Adaptive Plans

Adjust query plans at runtime
based upon current data

Adaptive Statistics

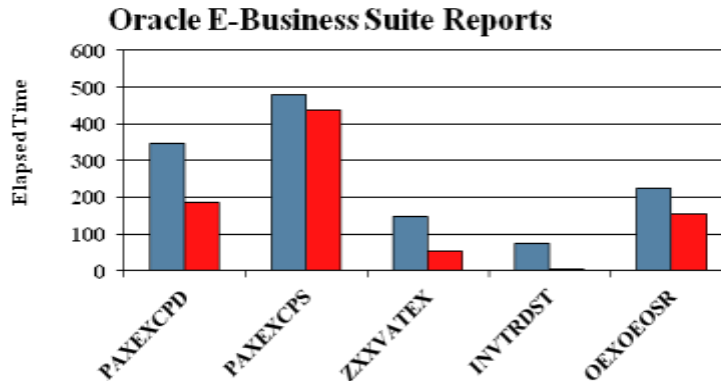
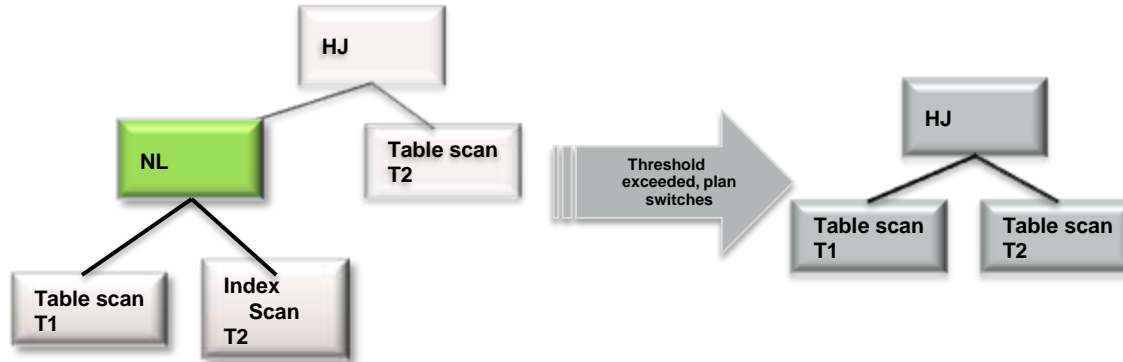
Adapt optimizer statistics at
runtime

“Learn” for future queries



Adaptive Execution Plans

Good SQL execution without intervention



- Plan decision deferred until runtime
- Final decision is based on statistics collected during execution
- If statistics prove to be out of range, sub-plans can be swapped
- Bad effects of skew eliminated

More Efficient Data Management

Partitioning Enhancements

- Enhanced Partition Maintenance operations
 - Online partition move
 - Partition maintenance operations on multiple partitions
 - Asynchronous global index maintenance for DROP and TRUNCATE
- Complete Reference Partitioning
 - Interval-Reference Partitioning
 - Cascade for TRUNCATE and EXCHANGE
- Partial global and local indexes

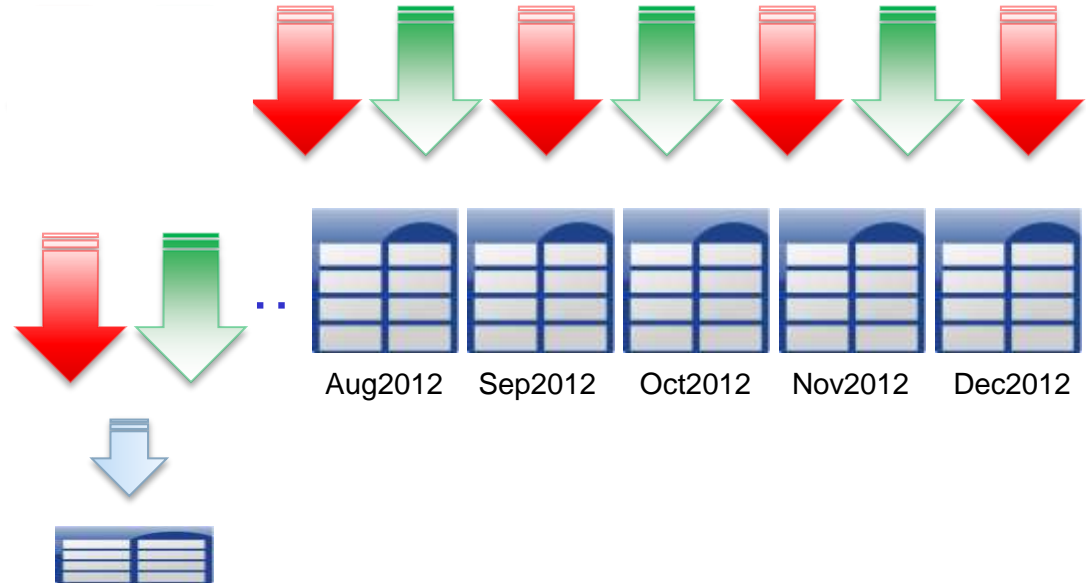
Enhanced Partition Maintenance Operations

Online Partition Move

Transparent MOVE
PARTITION ONLINE
operation

Concurrent DML and
Query

Index maintenance for
local and global indexes



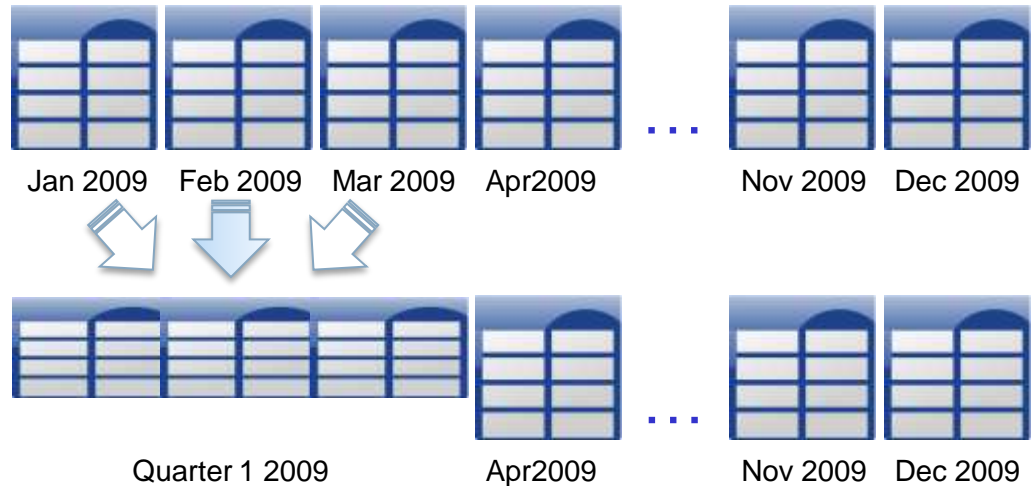
Enhanced Partition Maintenance Operations

“PMOP” on multiple partitions

Partition Maintenance on multiple partitions in a single operation

Full parallelism

Transparent maintenance of local and global indexes



```
ALTER TABLE orders
MERGE PARTITIONS Jan2009, Feb2009, Mar2009
INTO PARTITION Quarter1_2009 COMPRESS FOR ARCHIVE HIGH;
```

Automatic Data Optimization

Usage based data compression

01110101010010
10000100010101

Hot Data

10101010111010100110101
11000010100010110111010
10100101001001000010001
01010110100101101001110
00010100100101000010010
00010001010101110011010
10100101001001000010001

3X

Advanced Row Compression

Warm Data

1010101011101010011010111000010100
0101101110101010010100100100001000
1010101101001011010011100001010010
0101000010010000100010101011010010
1000010100100101001010111000010
1110010100100101001010110111011010

10X

Column Query Compression

Archive Data

10101010111010100110101110000101000101011
101010100101001001000010001010101101001011
010011100001010010010100001001000010001010
101010101110101001101011100001010001011011
101010101110101001101011100001011101011001

15X

Column Archive Compression

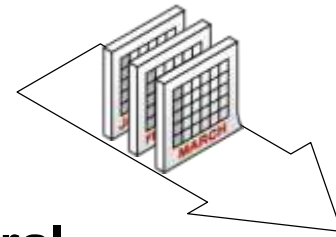
ORACLE

Data Redaction

Dyanamically Masking for Data Warehouses



Temporal Support



Transaction Time Temporal (Flashback Data Archive)

- Tracks transactional changes to a table over its lifetime
- Typically used for compliance and auditing
- Enables the users to see the data as it was at a point in time in the past

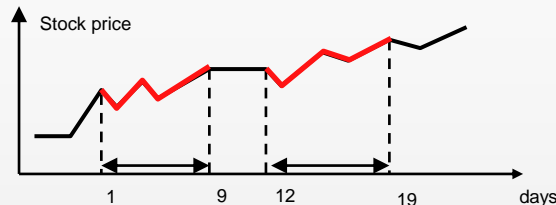
Valid Time Temporal

- Enables user to model & query data for “real world validity”
- Typically used for insurance policies, financial markets, trade data & future changes
- Users can model concepts such as the “Life time of an insurance policy”

SQL Pattern Matching

Find event A (“privilege revoked”) followed by 3 or more occurrences of event B (“attempted login”) within 1 minute

Find 10-day periods where a stock price has “double-bottomed”



SQL Pattern Matching provides expressive syntax and fast performance for pattern matching

New SQL construct: `MATCH_RECOGNIZE`

Define patterns using regular expression syntax

SQL Pattern Matching

Finding Double Bottom (W)

```
if (!q.isEmpty() && (next.isEmpty() || (gt(q, prev) && eq(q, next)))) {
    state = "E";
    return state;
}

if (q.isEmpty() || eq(q, prev)) {
    state = "F";
    return state;
}

return state;
}

private boolean eq(String a, String b) {
    if (a.isEmpty() || b.isEmpty()) {
        return false;
    }
    return a.equals(b);
}

private boolean gt(String a, String b) {
    if (a.isEmpty() || b.isEmpty()) {
        return false;
    }
    return Double.parseDouble(a) > Double.parseDouble(b);
}

private boolean lt(String a, String b) {
    if (a.isEmpty() || b.isEmpty()) {
        return false;
    }
    return Double.parseDouble(a) < Double.parseDouble(b);
}

public String getState() {
    return this.state;
}
```

250+ Lines of Java and PIG

```
SELECT first_x, last_z
FROM ticker MATCH_RECOGNIZE (
    PARTITION BY name ORDER BY time
    MEASURES FIRST(x.time) AS first_x,
              LAST(z.time) AS last_z
    ONE ROW PER MATCH
    PATTERN (X+ Y+ W+ Z+)
    DEFINE X AS (price < PREV(price)),
           Y AS (price > PREV(price)),
           W AS (price < PREV(price)),
           Z AS (price > PREV(price) AND
                z.time - FIRST(x.time) <= 7 ))
```

12 Lines of SQL

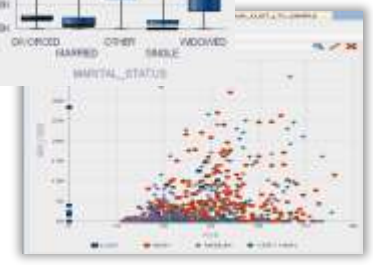
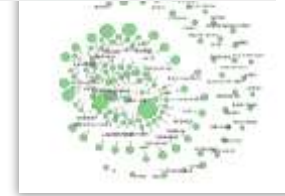
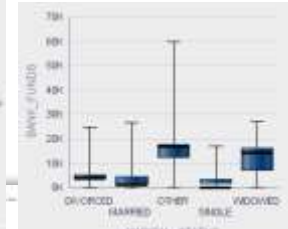
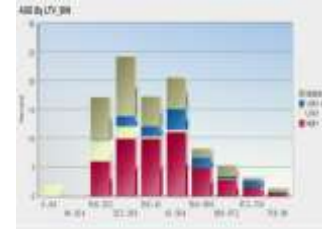
20x less code, 5x faster

Oracle Advanced Analytics

Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics

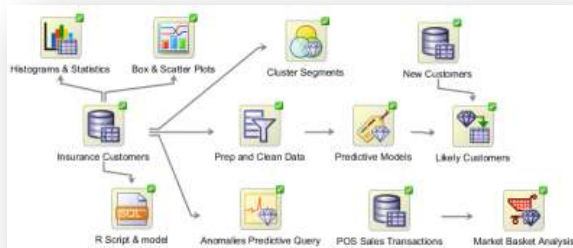
Key Features

- In-database data mining algorithms and open source R algorithms
- SQL, PL/SQL, R languages
- Scalable, parallel in-database execution
- Workflow GUI and IDEs
- Integrated component of Database
- Enables enterprise analytical applications



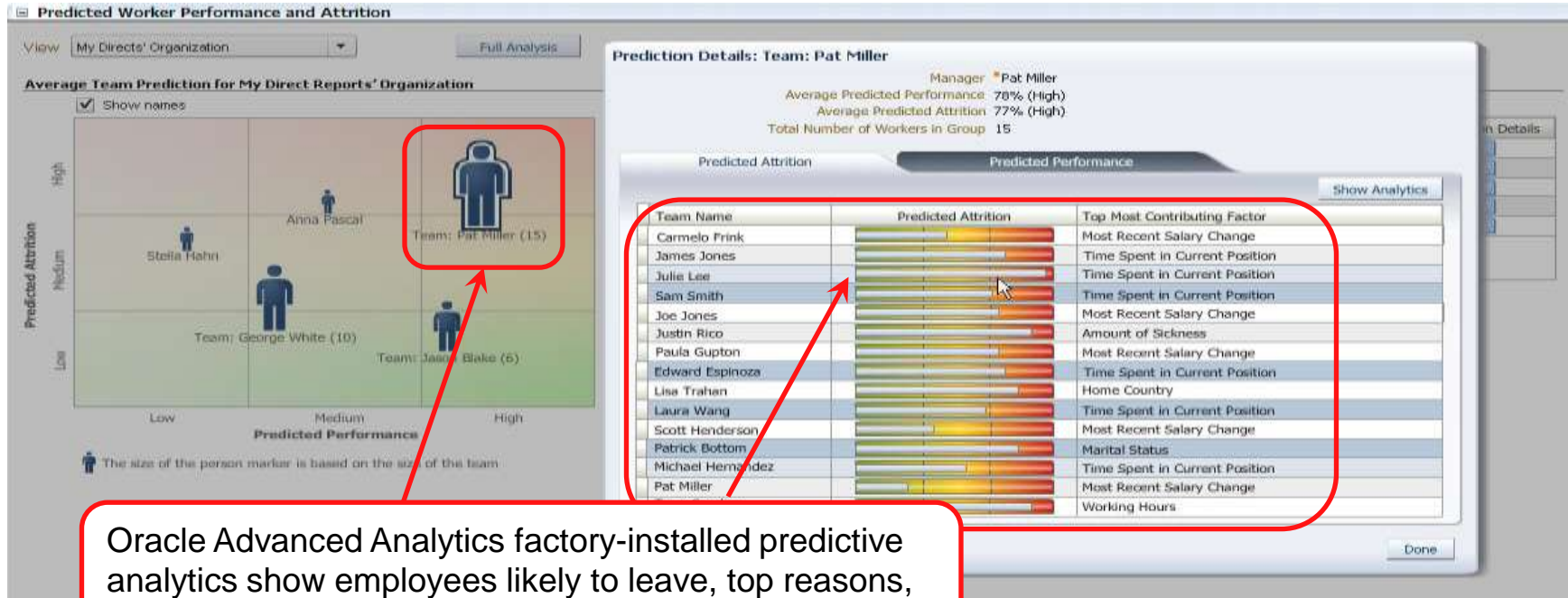
Why Oracle Advanced Analytics?

Differentiating Features



- ✓ Performance and Scalability
 - Leverages power and scalability of Oracle Database.
- ✓ Fastest Way to Deliver Enterprise Predictive Analytics Applications
 - Integrated with OBIEE and any application that uses SQL queries
- ✓ Lowest Total Costs of Ownership
 - No need for separate analytical servers

OAA in Fusion HCM Predictive Analytics



Oracle Advanced Analytics factory-installed predictive analytics show employees likely to leave, top reasons, expected performance and real-time "What if?" analysis

Sample of Big Data Use Cases Today

AUTOMOTIVE

Auto sensors
reporting
location,
problems



COMMUNICATIONS

Location-based
advertising



Retail / CPG

Sentiment analysis
Hot products
Optimized Marketing



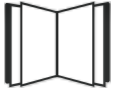
Games

Adjust to
player
behavior
In-Game Ads



EDUCATION & RESEARCH

Experiment
sensor analysis



HIGH TECH INDUSTRIES

Mfg quality
Warranty

**If you could have all the data you wanted,
what would you do differently?**



MEDIA/ ENTERTAINMENT

Viewers / advertising
effectiveness
Cross Sell



FINANCIAL SERVICES

Risk & portfolio analysis
New products



LIFE SCIENCES

Clinical trials
Genomics



UTILITIES

Smart Meter
analysis for
network
capacity,



LAW ENFORCEMENT & DEFENSE

Threat analysis -
social media
monitoring, photo
analysis



Oracle Big Data Platform

Oracle Big Data Appliance

Optimized for Hadoop, R, and NoSQL Processing

Hadoop

Open Source R

Oracle NoSQL Database

Applications

Oracle Big Data Connectors

Oracle Big Data Connectors

Oracle Data Integrator

Oracle Exadata "System of Record"

Optimized for DW/OLTP

Oracle Advanced Analytics
Data Warehouse
Oracle Database

In-Database Analytics

Oracle Exalytics

Optimized for Analytics & In-Memory Workloads

Oracle Enterprise Performance Management
Oracle Business Intelligence Applications
Oracle Business Intelligence Tools
Oracle Endeca Information Discovery

Stream

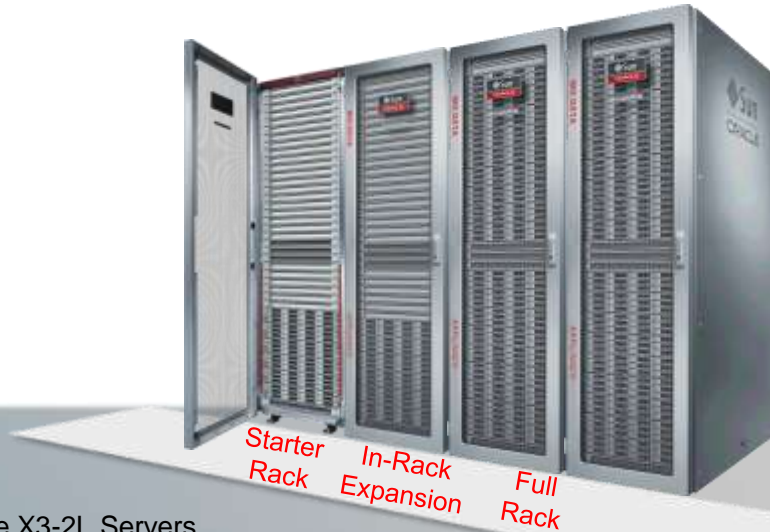
Acquire

Organize

Discover & Analyze

ORACLE

Big Data Appliance Product Family



- Starter Rack is a fully cabled and configured for growth with 6 servers
- In-Rack Expansion delivers 6 server modular expansion block
- Full Rack is 18 racks, extended with additional racks

- Sun Oracle X3-2L Servers
- Per node:
 - 2 sockets, 16 cores Intel Xeon
 - 64 GB Memory
 - 12 x 3 TB Disks

Big Data Appliance X3-2

Full Rack Only Configuration

Hardware

216 Intel® Xeon® Processors
864 GB total memory
64 GB per node
648TB total raw storage capacity
216 3TB 7200RPM Drives
40Gb/sec InfiniBand Network
10Gb/sec Data Center Connectivity

Software

Cloudera CDH
Cloudera Manager
Oracle Enterprise Manager Grid
Control Plug-In for BDA
NoSQL DB Community Edition
Open Source R



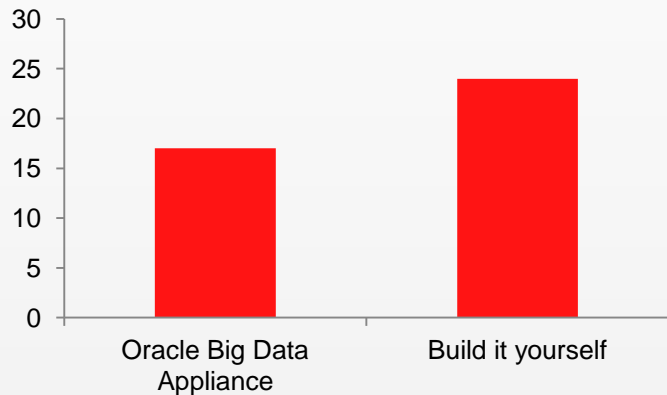
Why Oracle Big Data Appliance?



- Beats DIY Clusters on:
 - Initial Cost and Time to Value
 - Performance and Scalability
 - Support and maintenance effort
- Pre-configured with leading Hadoop Distribution
 - Proven at large scale
 - Contributors across all components for better support
- Better Integration with your Oracle ecosystem with:
 - High-performance connectivity to Exadata
 - Unified analytics API (SQL, R, MapReduce etc.)
 - Single Enterprise Manager Framework

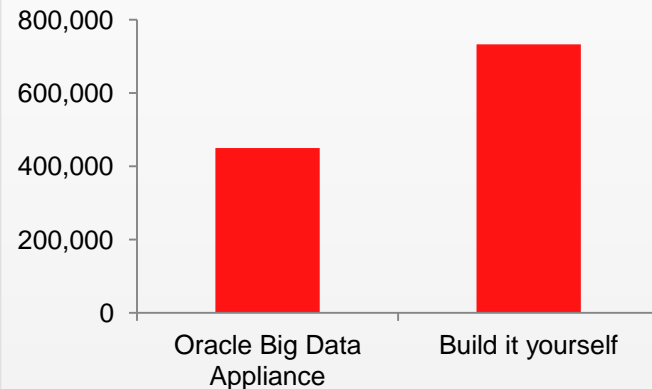
Engineered for Quicker Time to Value and Lower Cost

Implement 8 weeks faster ...



ESG believes that a "buy" versus "do-it-yourself" approach will yield roughly **one-third faster time-to-market** benefit improvement...

... for \$283K less money



[...] nearly **40% cost savings** versus IT architecting, designing, procuring, configuring, and implementing its own big data infrastructure.

Enterprise Manager Plug-In

Full Hardware
Management for BDA

See all software status
from Enterprise Manager
at a glance

Tight Integration with
Cloudera CM 4 for
Hadoop Management

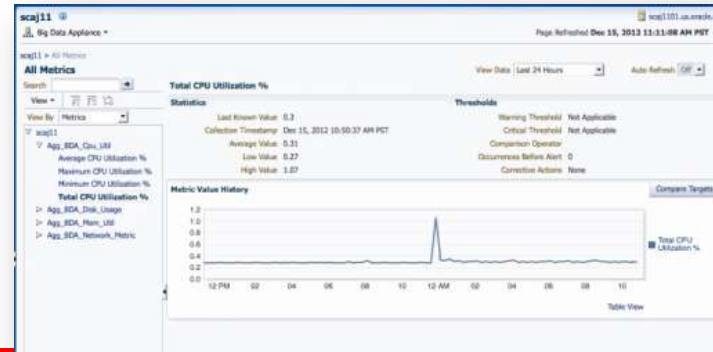
Infiniband Network



Host Level Metrics



CPU Utilization



ORACLE

Big Data Connectors

Optimized integration of Hadoop with Oracle Database and Oracle Exadata

- Oracle Loader for Hadoop
 - Oracle SQL Connector for Hadoop Distributed File System (HDFS)
 - Oracle Data Integrator Application Adapter for Hadoop
 - Oracle R Connector for Hadoop
-
- Does not require Big Data Appliance – can be licensed for Hadoop running on non-Oracle hardware

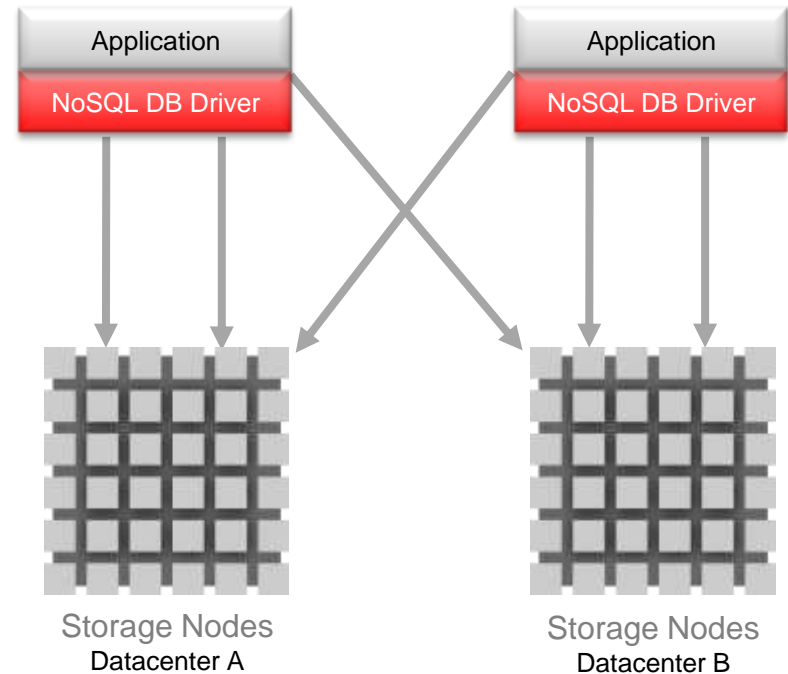


Oracle NoSQL Database

Scalable, Highly Available, Key-Value Database

Key Features

- Simple Key-Value Data Model
- Horizontally Scalable
- Highly Available
- Simple administration
- ACID Transactions at scale
- Transparent load balancing
- Elastic Configuration
- Commercial grade software and support



Oracle NoSQL Database: Use Cases

Web-Scale Transaction Processing

- High velocity, volume, variety, Low information density data capture

- Apps: Web browsing, Shopping Carts, CDR processing, Sensor data capture

Last-Mile Content Delivery

- Guaranteed low latency lookups for end-customers

- Apps: Advertising, Product Recommendations, Catalogs, Social Media

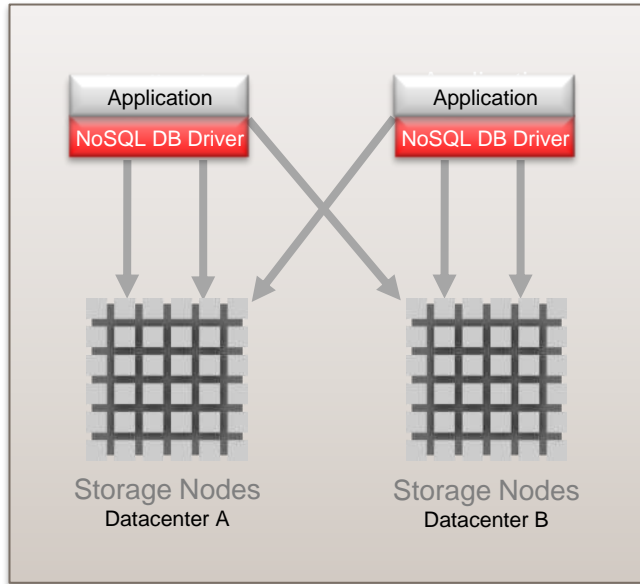
Real-Time Event Processing

- Real time events trigger rule execution in Oracle Event Processing
- OEP performs low latency lookups in NoSQL DB for additional data

- Apps: Medical Monitoring, Factory Automation, Oil & Gas, Geo-location

Why Oracle NoSQL Database?

Differentiating Features



- ✓ Flexibility
 - Simple Key-Value data model
 - Configurable ACID Transactions at scale
- ✓ Ease of use
 - Smart Topology
- ✓ Integration with the Oracle technology stack

Oracle for Data Warehousing and Big Data

- Today's data warehouses grow and increase in relevance
 - New technology in Oracle Database 12c
- The evolution of data warehousing is underway
 - New Big Data technologies are blending with current data warehouses



For more information

- **Big Data**

- <http://www.oracle.com/us/technologies/big-data/index.html>
- <http://bigdata.us.oracle.com>

- **Focus on Data Warehousing**

- <http://www.oracle.com/technetwork/database/bi-datawarehousing/index.html>

Hardware and Software

ORACLE

Engineered to Work Together

ORACLE®