Build versus Buy:
Big Data for the Enterprise

Lajos Sárecz

Principal Sales Consultant

Oracle Hungary

Agenda

- Architecting Big Data

- Big Data Products

- Build versus Buy

- Connect Existing Infra with Big Data

- Summary

Architecting Big Data

ORACLE

Sample of Big Data Use Cases Today

AUTOMOTIVE
Auto sensors reporting location, problems

HIGH TECHNOLOGY AND INDUSTRIAL MFG.
Mfg quality
Warranty analysis

OIL & GAS
Drilling exploration sensor analysis

Clinical trials

MEDIA
People & career

EDUCATION & RESEARCH
Experiment sensor analysis
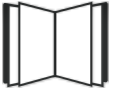
HEALTH CARE
Patient sensors, monitoring, EHRs
Quality of care

LAW ENFORCEMENT & DEFENSE
Threat analysis - social media monitoring, photo analysis

What is the main difference in this data?

Volume, Velocity, Variety

These Characteristics Challenge your Existing Architecture

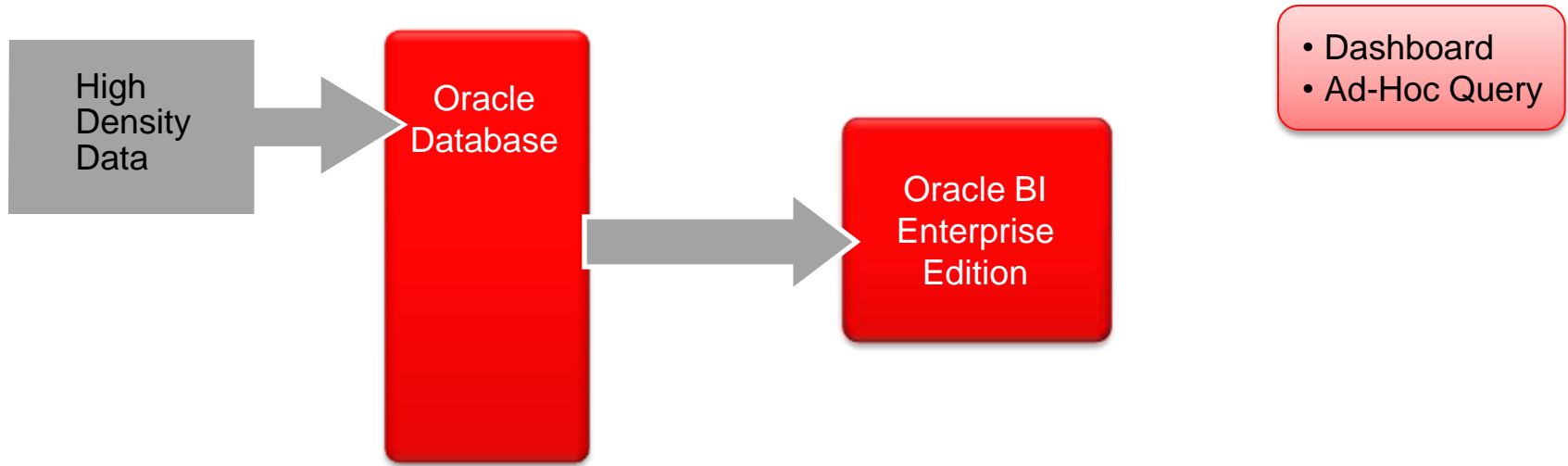**ORACLE**

Building your big data architecture

Gradually Extending your Existing Architecture for Big Data:

- Step 1: Further Analyze Current Data
- Step 2: Architect for Data Variety and Volume
- Step 3: Architect for Data Velocity
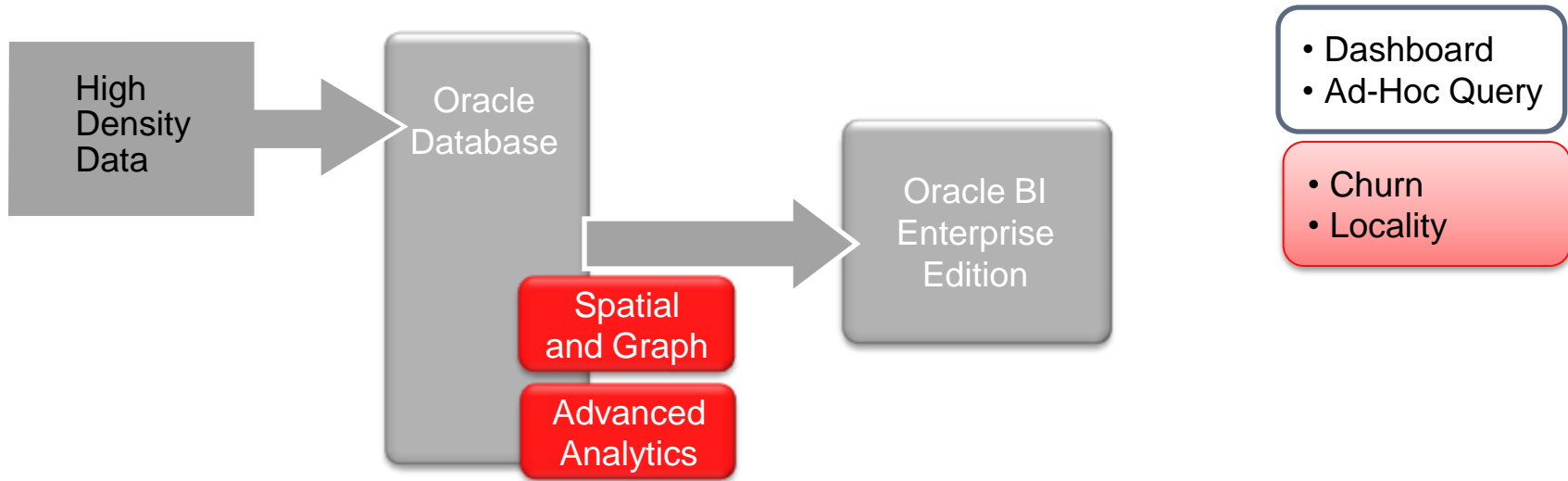- Step 4: Discover New Patterns

Increase → **Business Value**

Step 0: Data Warehouse Foundation

```
High
Density
Data   ──▶   Oracle
              Database   ──▶   Oracle BI
                               Enterprise
                               Edition
```

- Dashboard
- Ad-Hoc Query

Acquire | Organize | Analyze | Decide ──▶

Step 1: Deep Analysis of Current Data

High Density Data

Oracle Database

Spatial and Graph

Advanced Analytics

Oracle BI Enterprise Edition

- Dashboard
- Ad-Hoc Query

- Churn
- Locality

Acquire — Organize — Analyze — Decide

ORACLE

Step 2: Architect for Volume and Variety

High Density Data

Low Density variable Data

Hadoop

Aggregate Pre-Analyze

Oracle Database

Spatial and Graph

Advanced Analytics

Oracle BI Enterprise Edition

- Dashboard
- Ad-Hoc Query

- Churn
- Locality

- Relationship
- Comments

Acquire | Organize | Analyze | Decide
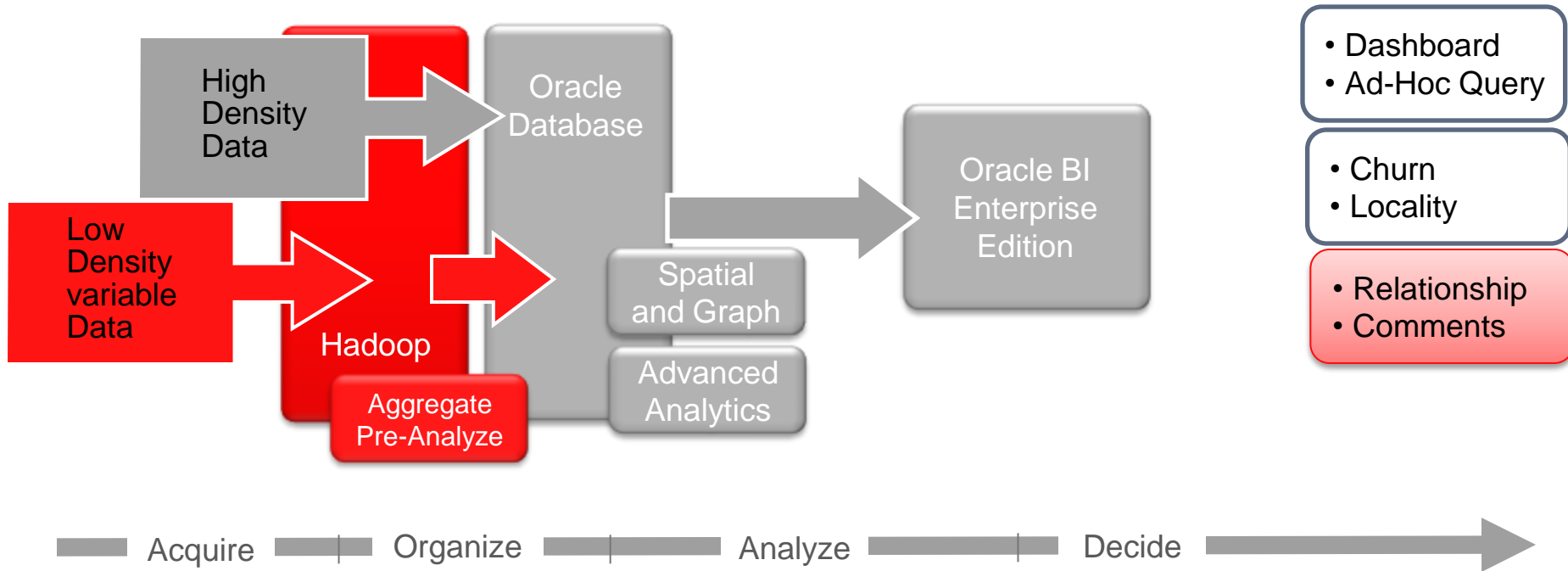
ORACLE

Step 3: Architect for Velocity
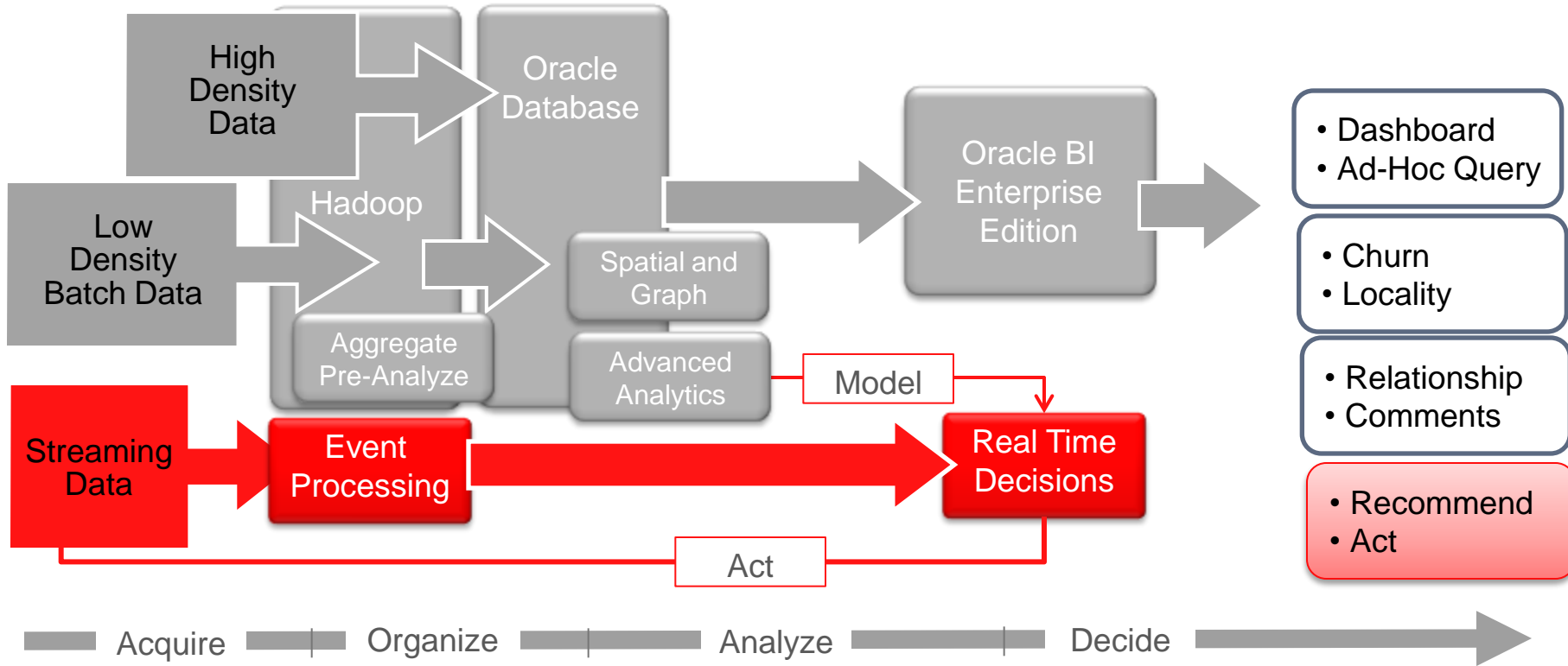
ORACLE

Step 4: Discover New Information

Endeca Information Discovery

High Density Data

Oracle Database

Oracle BI Enterprise Edition

Hadoop

Low Density Batch Data

Spatial and Graph

Aggregate Pre-Analyze

Advanced Analytics

Model

Streaming Data

Event Processing

Real Time Decisions

Act

Acquire | Organize | Analyze | Decide

- Dashboard
- Ad-Hoc Query

- Churn
- Locality

- Relationship
- Comments

- Recommend
- Act

- Discover

ORACLE

Oracle Engineered Systems Simplify Big Data

Oracle Big Data Appliance

Oracle Exadata

Oracle Exalytics

InfiniBand

InfiniBand

- Dashboard
- Ad-Hoc Query

- Churn
- Locality

- Relationship
- Comments

- Recommend
- Act

- Discover

Acquire   Organize   Analyze   Decide

ORACLE

Big Data Products

ORACLE

Big Data Appliance X3-2

Sun Oracle X3-2L Servers with per server:

- 2 * 8 Core Intel Xeon E5 Processors
- 64 GB Memory
- 36TB Disk space

Totals per Full Rack:

- 288 Processor Cores
- 1152 GB of Memory
- 648TB Available Disk space

ORACLE

Big Data Appliance Software Stack

Integrated Software:

- Oracle Linux 5.8 with UEK
- Cloudera CDH 4.2 & Cloudera Manager 4.5
- Big Data Appliance Enterprise Manager Plug-In
- Oracle R Distribution

All integrated software is supported as part of Premier Support for Systems and Premier Support for Operating Systems

Optional Software:

- Oracle NoSQL Database 2.x
- Oracle Big Data Connectors 2.x

ORACLE

Big Data Appliance Product Family

Starter Rack · In-Rack Expansion · Full Rack

- Starter Rack is a fully cabled and configured for growth with 6 servers
- In-Rack Expansion delivers 6 server modular expansion block
- Full Rack delivers optimal blend of capacity and expansion options
- Grow by adding rack – up to 18 racks without additional switches

ORACLE

Big Data Appliance X3-2 Starter Rack

Start and grow in increments of six servers

- 6 Nodes fully cabled in Starter Rack
  - 96 Intel® Xeon® E5 Processors
  - 384 GB total memory
  - 216TB total raw storage capacity
- 6 Nodes In-Rack Expansion added in-rack
  - 96 Intel® Xeon® E5 Processors
  - 384 GB total memory
 - 216TB total raw storage capacity

ORACLE

Why Oracle Big Data Appliance?

- Beats DIY Clusters on:
  - Initial Cost and Time to Value
  - Performance and Scalability
- 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.)
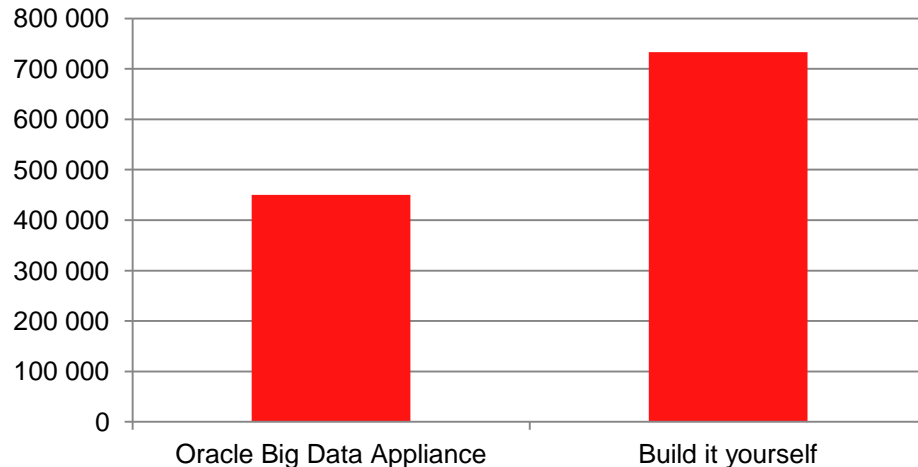 - Single Enterprise Manager Framework

ORACLE

Engineered for Quicker Time to Value at Lower Cost

Time to Market (Weeks)

```
30
25
20
15
10
 5
 0
         Oracle Big Data Appliance    Build it yourself
```
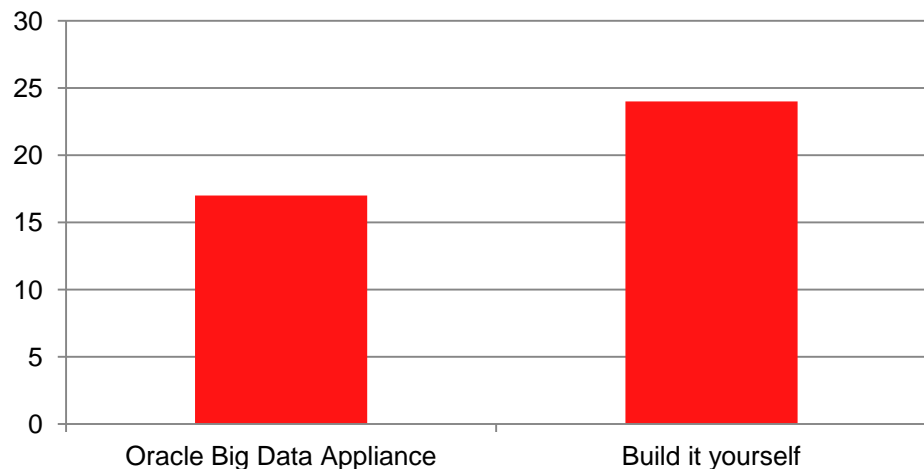
ESG believes that a "buy" versus "do-it-yourself" approach will yield roughly one-third faster time-to-market benefit improvement...

Cost: Initial Infrastructure/Tasks

```
800 000
700 000
600 000
500 000
400 000
300 000
200 000
100 000
      0
         Oracle Big Data Appliance    Build it yourself
```

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

ORACLE

 http://www.oracle.com/us/corporate/analystreports/industries/esg-big-data-wp-1914112.pdf

Engineered for Performance

Compared with a DIY Cluster

- Configured for exceptional performance on delivery

- 6x faster than custom 20-node Hadoop cluster for large batch transformation jobs

- Engineering done by Oracle and Cloudera:

  - OS and File System Tuning

  - Java Virtual Machine Tuning

  - Hadoop Configuration and Setup

Time (hours)

10

5

0

6x

Big Data Appliance

DIY Hadoop Cluster
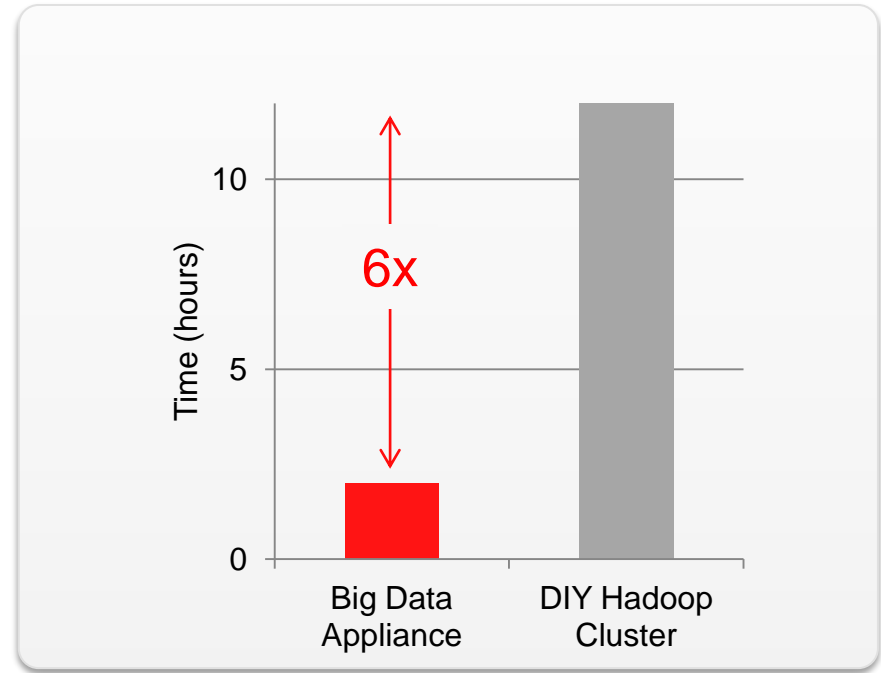
ORACLE

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

Analyze Data across your Oracle Systems

SQL Analytics on ALL data

- Expand the data pool for analytics leveraging Hadoop

- Stream Hadoop resident data through Big Data Connectors for SQL processing

- Use the full power of Oracle SQL on all data

- Or use Oracle Loader for Hadoop to integrate data in Oracle Database

SQL

Hadoop

Oracle Database

IB
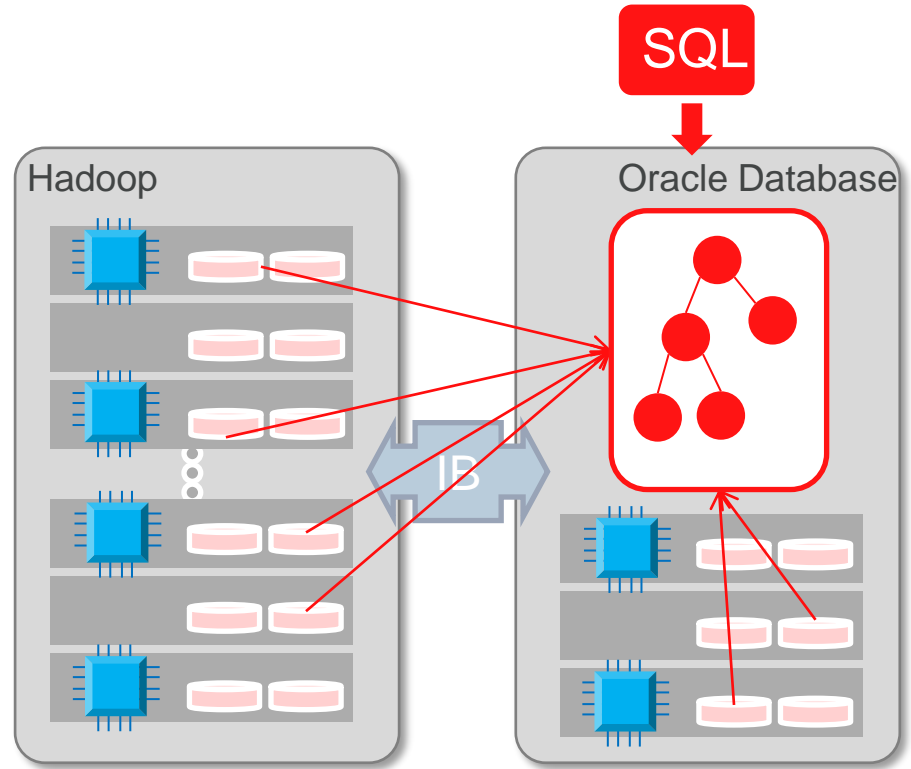
ORACLE

# Analyze Data across your Oracle Systems

R Analytics on ALL data

- Expand the data pool for analytics leveraging Hadoop

- Improve scalability and performance for R without changes to your programs

- Dynamically leverage Hadoop through Big Data Connectors to execute R analytics
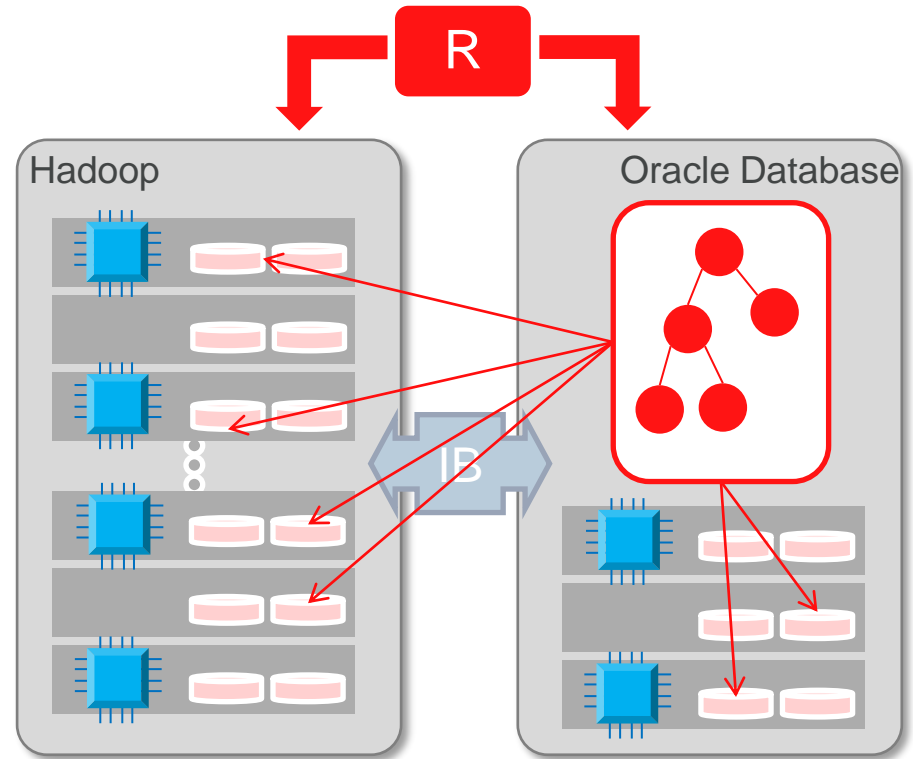
ORACLE

Oracle Data Integrator

Simplify Map Reduce

- Automatically generates MapReduce code

- High performance loads into Data Warehouse leveraging both OLH and OSCH

- Manages the process across platforms
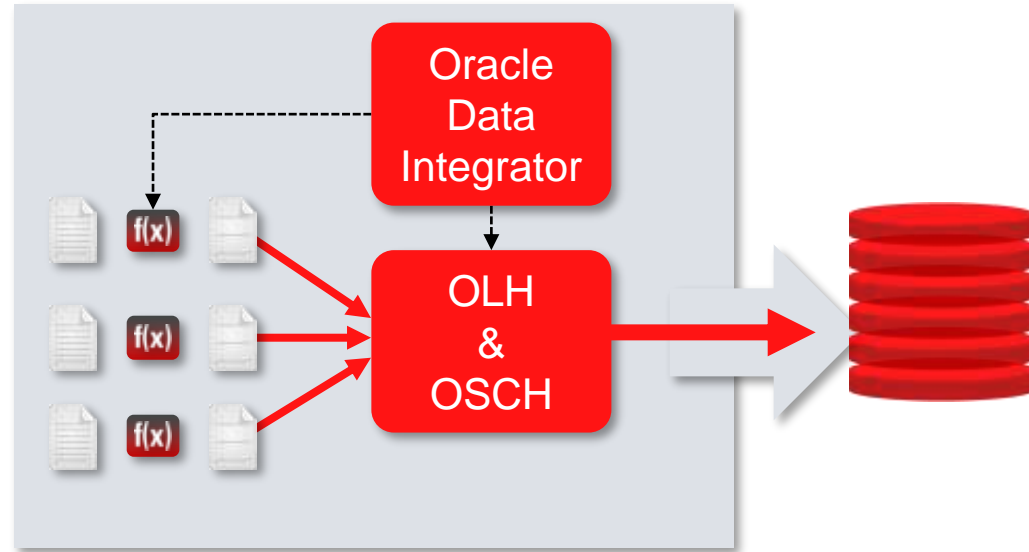
ORACLE

Oracle Big Data Solution
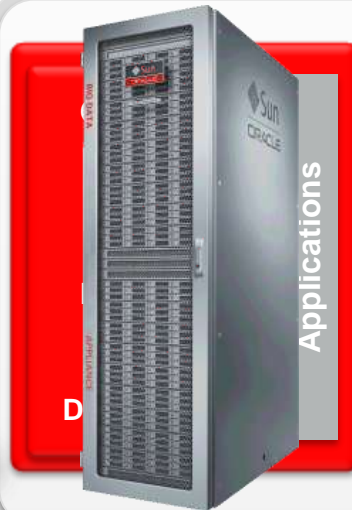

Decide

Oracle Real-Time Decisions

Oracle BI Enterprise Edition

Endeca Information Discovery

- Complete
- Integrated
- Scalable

Applications

In-Database Analytics

Stream

Acquire – Organize – Analyze

 | Insert Information Protection Policy Classification from Slide 8