ORACLE®

An Integrated Analytics & Big Data Infrastructure
September 21, 2012

Robert Stackowiak, Vice President Data Systems Architecture
Oracle Enterprise Solutions Group

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.

Common Data Warehousing Areas of Focus

Typical Structured Data Analysis Today

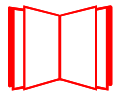
AUTOMOTIVE
- Mfg cost analysis
- Cost of service / warranty analysis

COMMUNICATIONS
- Logistics optimization
- Network analysis

CONSUMER PACKAGED GOODS
- Supplier & channels analysis
- Consumer trends

FINANCIAL SERVICES
- Risk & portfolio analysis
- Customer analysis

EDUCATION & RESEARCH
- Cost of facilities & staff analysis
- Academic & alumni profile



HIGH TECHNOLOGY / INDUSTRIAL MFG.
- Customer & distributor analysis
- Mfg cost analysis

LIFE SCIENCES
- Clinical trials
- Cost of research & production

MEDIA/ ENTERTAINMENT
- Viewer & channels analysis
- Venue optimization

ON-LINE SERVICES
- Customer analysis & site statistics
- Cross-sell / up-sell

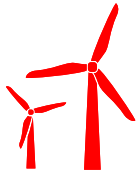
HEALTH CARE
- Cost of care
- Quality of care
- Staffing analysis

OIL & GAS
- Drilling exploration costs & logistics optimization

RETAIL
- Market basket analysis
- Supply chain optimization
- Real estate optimization

TRAVEL & TRANSPORTATION
- Equipment & crew logistics & routing optimization
- Customer analysis

UTILITIES
- Equipment logistics optimization
- Customer analysis
- Grid cost of delivery



LAW ENFORCEMENT & DEFENSE
- Logistics optimization
- Crime statistics analysis

Challenged by: **Growing Data Volume & More Complex Analysis Needs**

Sources for the Data are Growing

- 383+ Million Twitter accounts (100m+ tweeting)
- 835+ Million Facebook subscribers
- 1.2+ Billion Mobile Web users
- Sensors everywhere

Structured Data & “Big Data”

Exhibit 25: Structured Data Example

Name	Thread pitch (mm)	Minor diameter tolerance	Nominal diameter (mm)	Head shape	Price for 50 screws	Available at factory outlet?	Number in stock	Flat or Phillips head?
M4	0.7	4g	4	Pan	$10.08	Yes	276	Flat
M5	0.8	4g	5	Round	$13.89	Yes	183	Both
M6	1	5g	6	Button	$10.42	Yes	1043	Flat
M8	1.25	5g	8	Pan	$11.98	No	298	Phillips
M10	1.5	6g	10	Round	$16.74	Yes	488	Phillips
M12	1.75	7g	12	Pan	$18.26	No	998	Flat
M14	2	7g	14	Round	$21.19	No	235	Phillips
M16	2	8g	16	Button	$23.57	Yes	292	Both
M18	2.1	8g	18	Button	$25.87	No	664	Both
M20	2.4	8g	20	Pan	$29.09	Yes	486	Both
M24	2.55	9g	24	Round	$33.01	Yes	982	Phillips
M28	2.7	10g	28	Button	$35.66	No	1067	Phillips
M36	3.2	12g	36	Pan	$41.32	No	434	Both
M50	4.5	15g	50	Pan	$44.72	No	740	Flat

Exhibit 26: Quasi-Structured Data Example

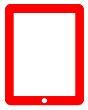
Structured data from applications.

Semi-structured “Big Data” from social media and logs, sensors, feeds, etc.

Big Data Fills Out the Complete Picture

AUTOMOTIVE
Auto sensors reporting location, problems

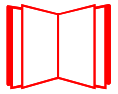
COMMUNICATIONS
Location-based advertising



CONSUMER PACKAGED GOODS
Sentiment analysis of what's hot, problems

FINANCIAL SERVICES
Risk & portfolio analysis

EDUCATION & RESEARCH
Experiment sensor analysis



HIGH TECHNOLOGY / INDUSTRIAL MFG.
Mfg quality
Warranty analysis

LIFE SCIENCES
Clinical trials
Genomics

MEDIA/ ENTERTAINMENT
Viewers / advertising effectiveness

ON-LINE SERVICES / SOCIAL MEDIA
People & career matching
Web-site optimization

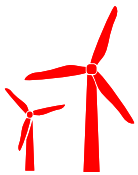
HEALTH CARE
Patient sensors, monitoring, EHRs
Quality of care

OIL & GAS
Drilling exploration sensor analysis

RETAIL
Consumer sentiment
Optimized sales & marketing

TRAVEL & TRANSPORTATION
Sensor analysis for optimal traffic flows
Customer sentiment

UTILITIES
Smart Meter analysis

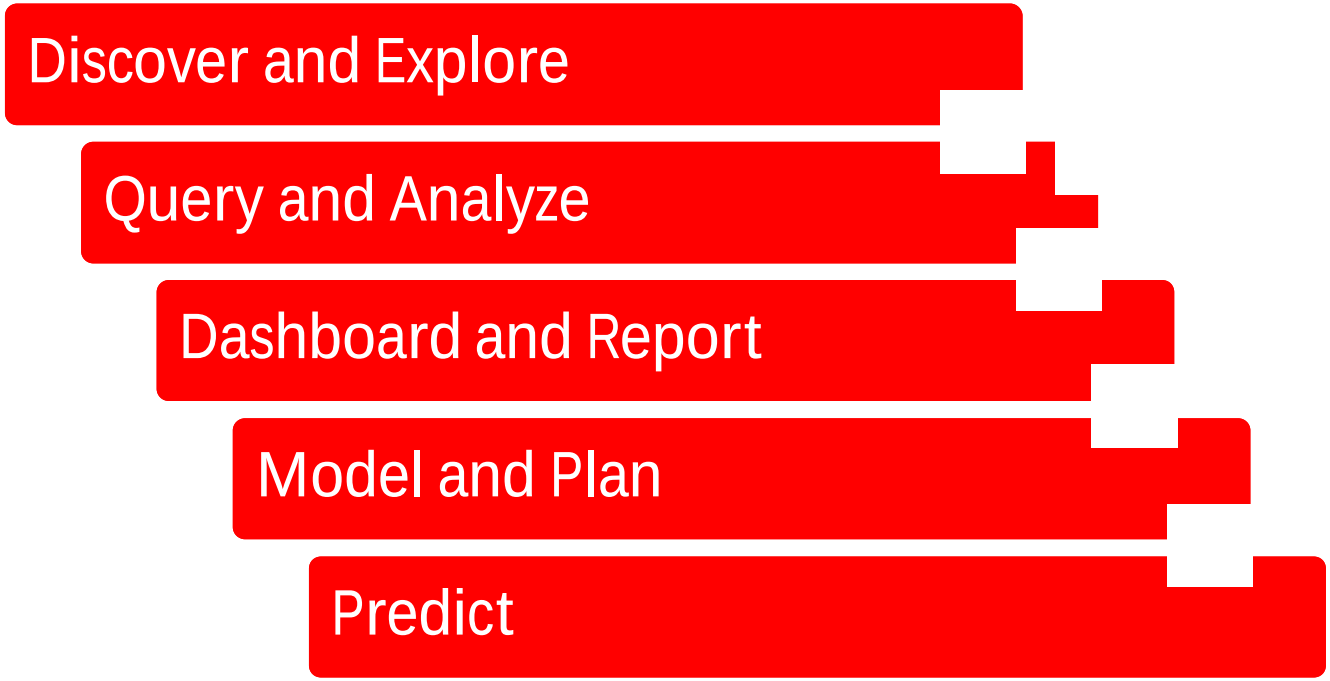


LAW ENFORCEMENT & DEFENSE
Threat analysis - social media monitoring, photo analysis

Challenged by: **Data Volume, Velocity, Variety in finding Value**

Typical Stages in Analytics

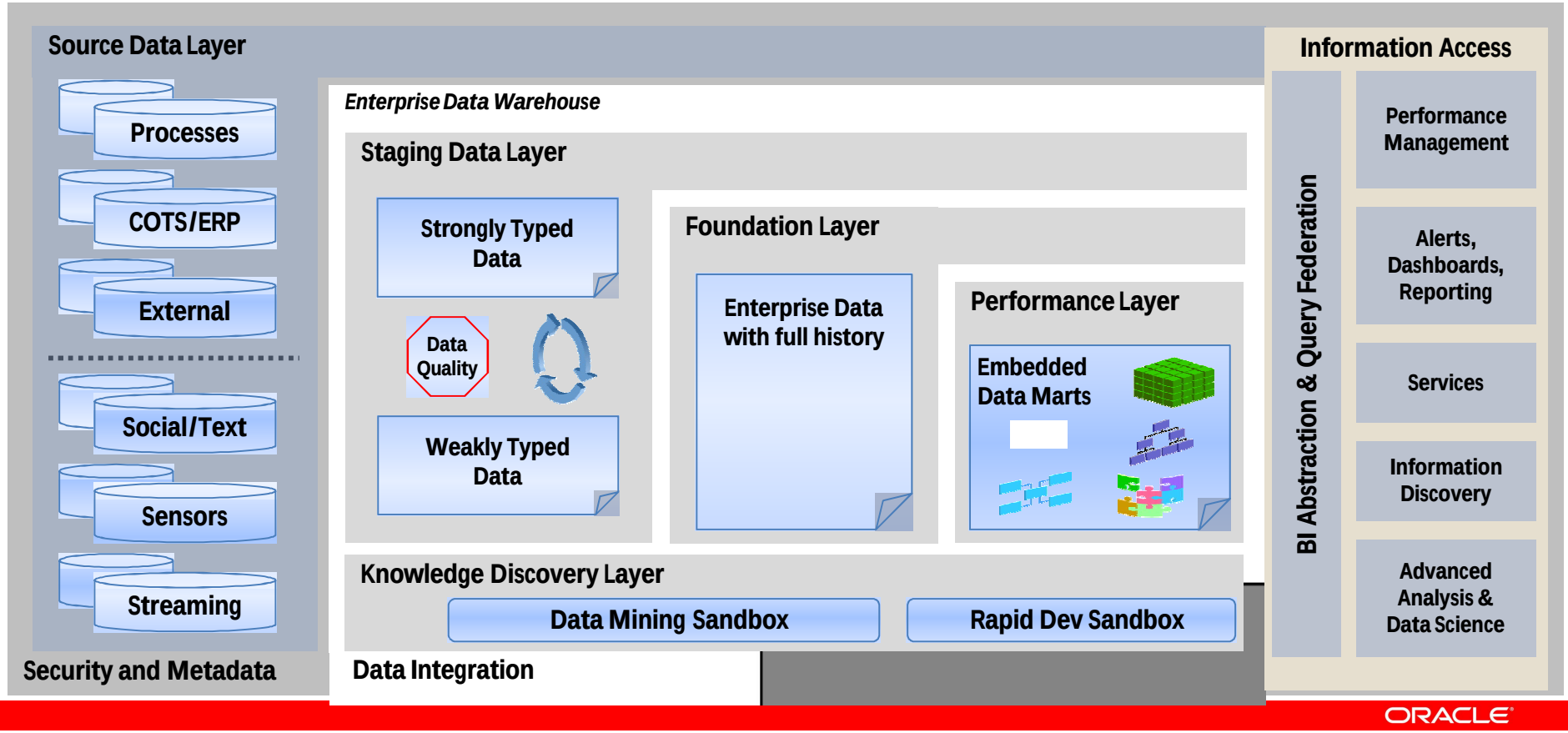
Choosing the Right Solutions for Right Data Needs



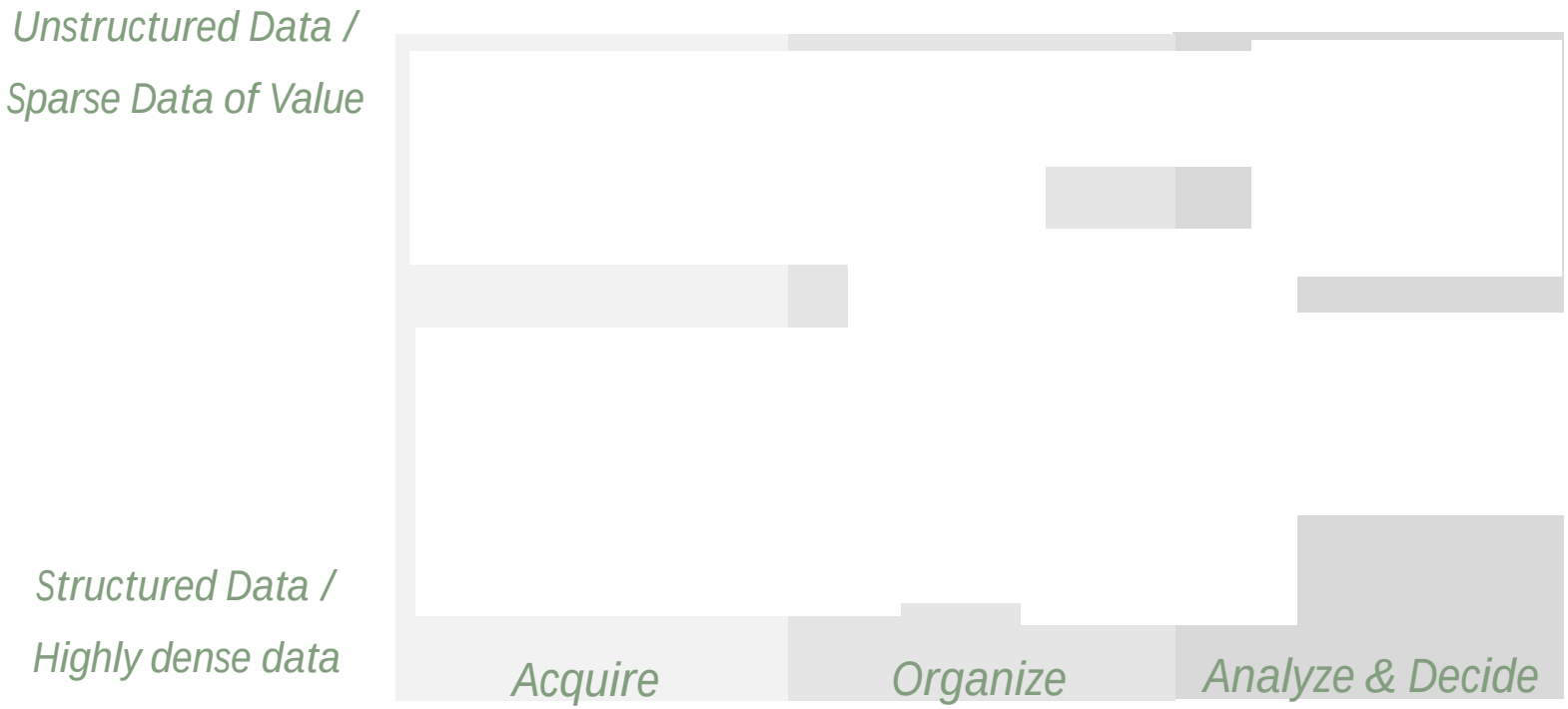
Challenges & Strategies

CHALLENGES	STRATEGIES
• Fragmented Solutions	• Specialized but integrated data stores and tools
• Difficulty of Self-Service BI	• Flexible, guided, automated, easy-to-use tools, data discovery
• Data Not Current	• Solutions for Just-in-Time well-understood data
• Time to ROI / Development Time	• Horizontal and industry pre-built solutions, appliance-like solutions & Cloud solutions
• Rapidly Growing Diverse Data & User Communities	• Enterprise class solutions serving 1000s of users optimized for diverse workloads and providing petabytes of data
• Deployment Manageability, Security & Expense	• Pre-integrated solutions that are centrally managed with advanced security / governance; Consolidation where possible to reduce platform footprint space & power

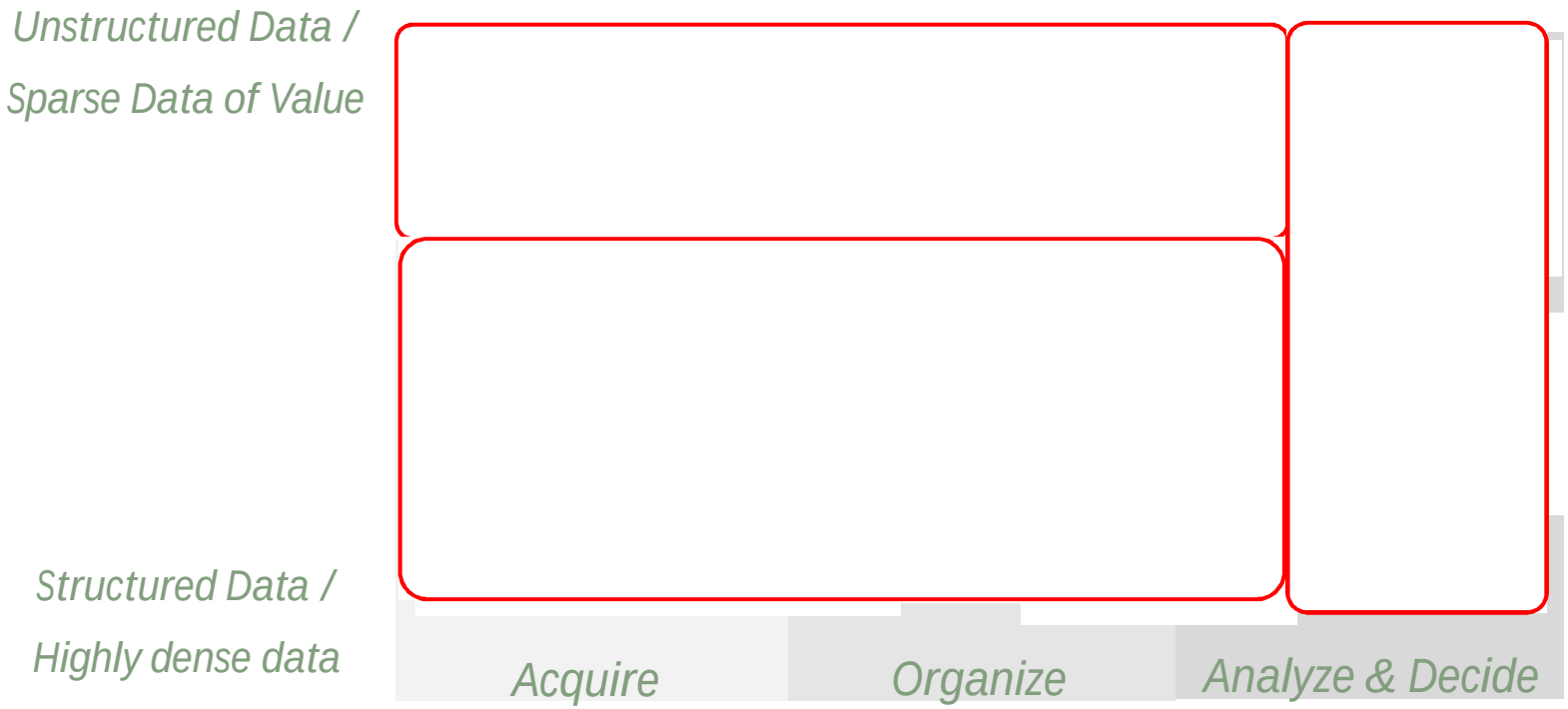
An Information Architecture that includes Big Data



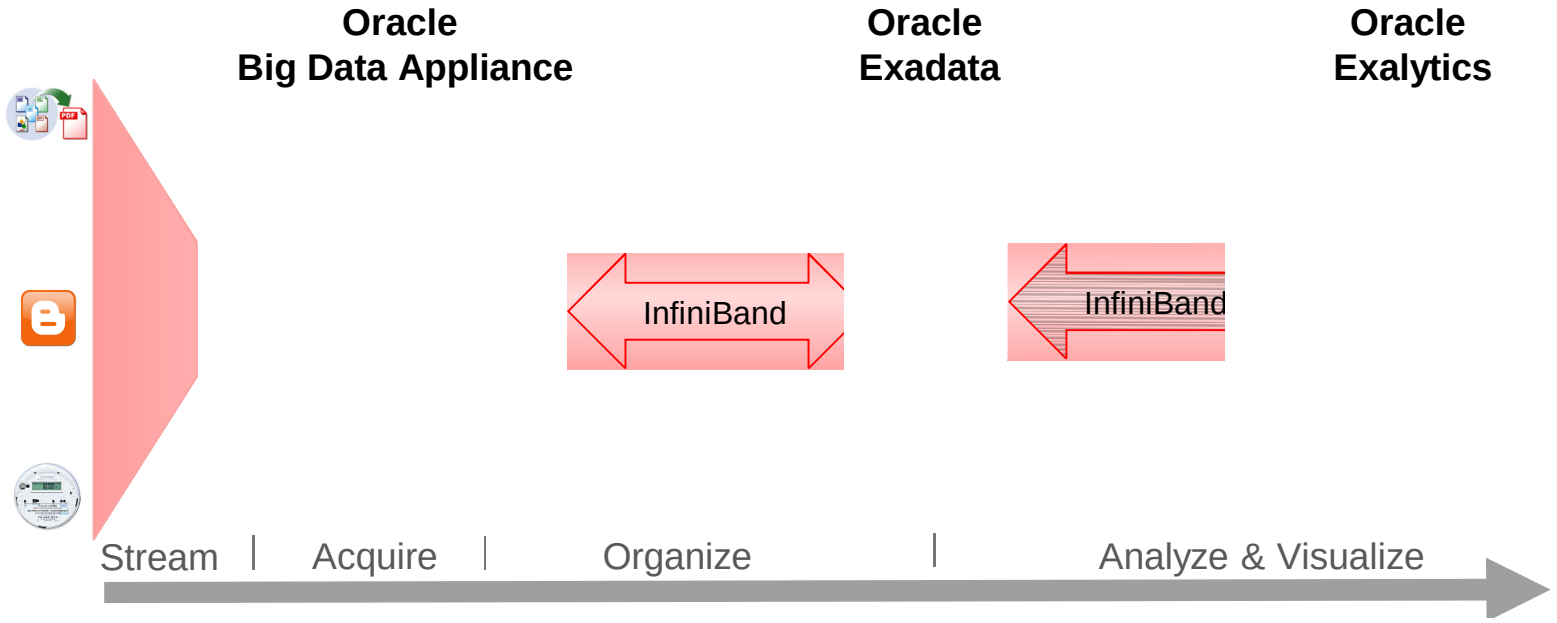
Oracle Analytics Software Components…



… & Engineered Systems



Oracle's Analytics Platforms



Big Data Appliance

Big Data for the Enterprise

- Foundation Software:
  - Oracle Linux
  - Oracle Java VM
  - Cloudera Apache Hadoop Distribution
  - Cloudera Manager
  - Oracle NoSQL Database Community Edition

- Application Software:
  - Oracle NoSQL Database Enterprise Edition – New
  - Oracle Big Data Connectors - New
    - Oracle Loader for Hadoop
    - Oracle Direct Connector for HDFS
    - Oracle Data Integrator Application Adapter for Hadoop
    - Oracle R Connector for Hadoop

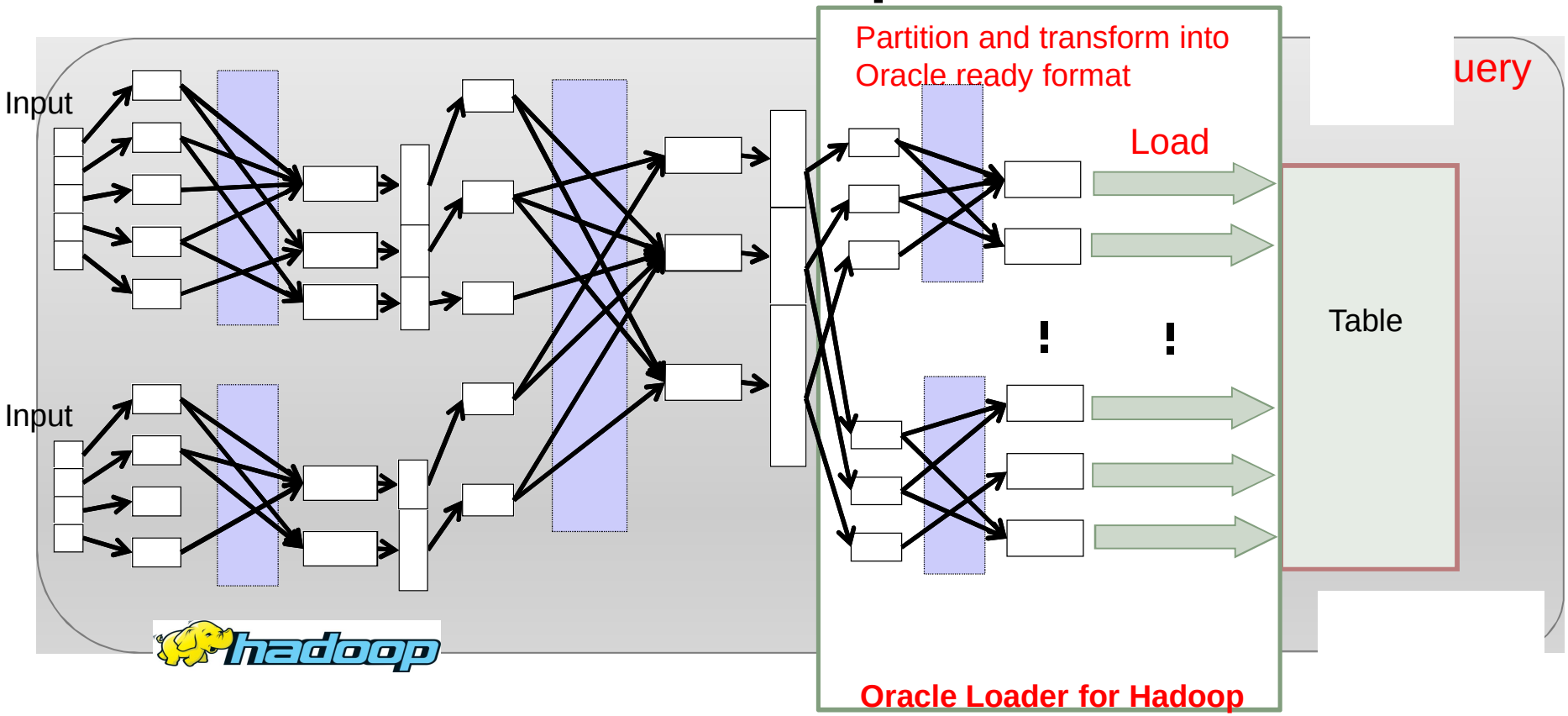
18 Sun X4270 M2 Servers

- 48 GB memory per node = 864 GB memory
- 12 Intel cores per node = 216 cores
- 36 TB storage per node = 648 TB storage

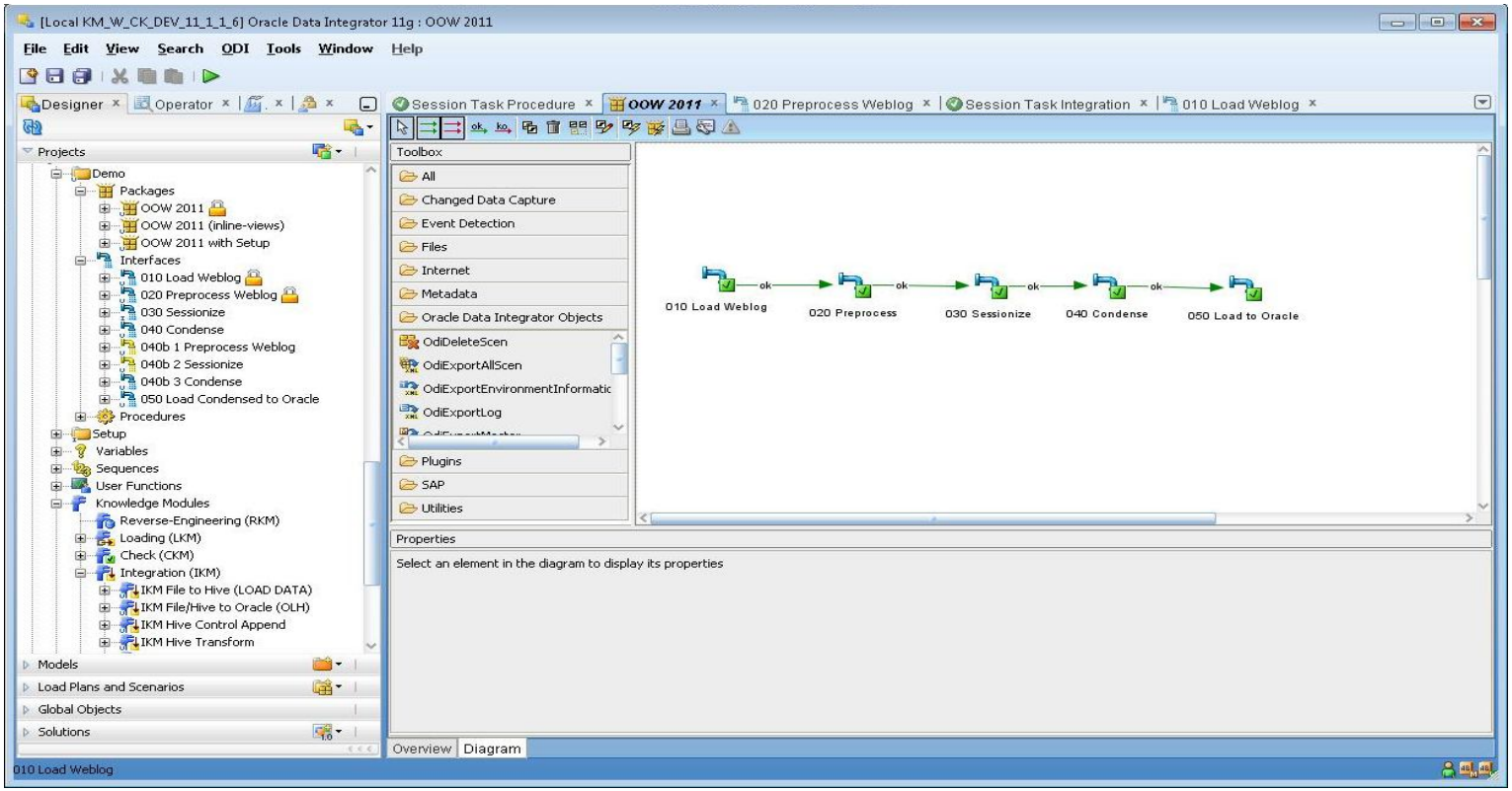
40 Gb/sec InfiniBand

10 Gb/sec Ethernet

Oracle Loader for Hadoop



Oracle Data Integrator & OLH

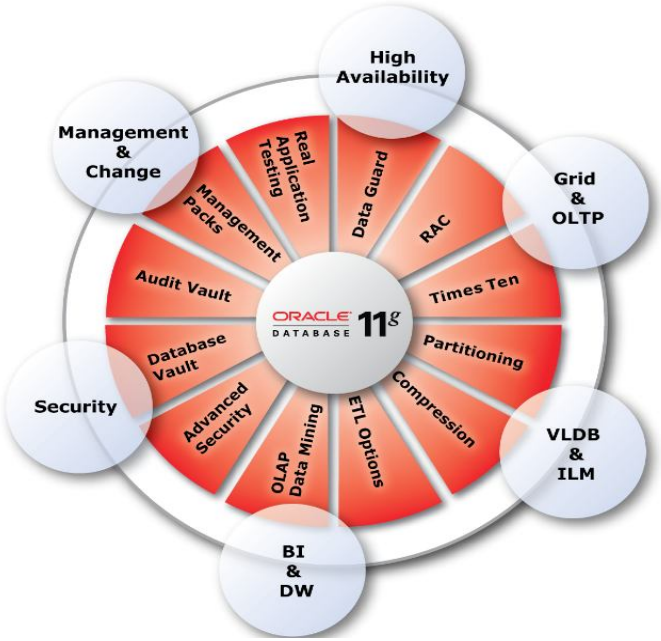


Oracle R Connector for Hadoop

- R package that provides an interface between the local R environment, Oracle Database, and Hadoop
- Using simple R functions, copy data between R memory, the local file system, Oracle Database, HDFS
- Schedule R programs to execute as Hadoop MapReduce jobs - return the results to any of the locations

Oracle Database 11*g* Data Warehousing

The Leading Database for Data Warehousing



Key Data Warehousing Capabilities

- Flexible Model Deployment
- Embedded Analytics
  - Advanced Analytics (R & Data Mining)
  - OLAP
- Single Point of Management
- Secure
- 24X7 Availability
- Optimal Storage Management
- Scaled to petabytes & large business analyst communities

Exadata Hardware Architecture

Database Grid

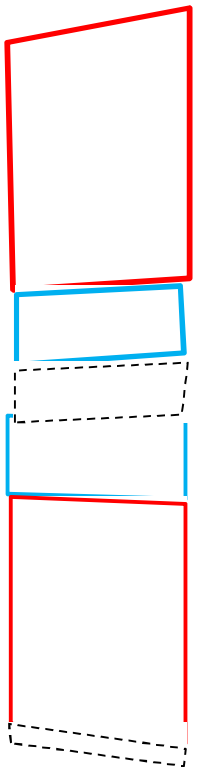
- 8 x Dual-processor x64 database servers

OR

- 2 x Eight-processor x64 database servers

InfiniBand Network

- 3 x 36-port 40Gb/s switches
- Unified server & storage network

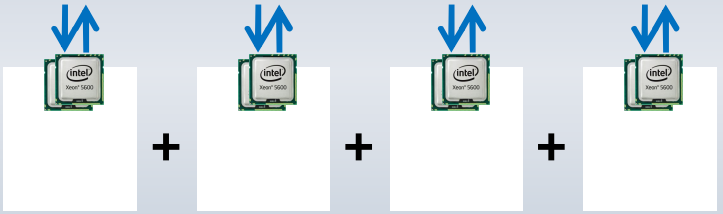


Intelligent Storage Grid

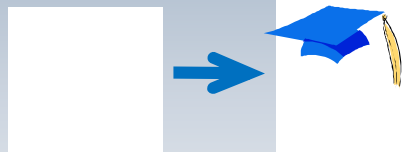
- 14 High-performance low-cost storage servers
- 100 TB High Performance disks, or 504 TB High Capacity disks
- 22.4 TB PCI Flash
- Intelligent Storage Server Software

Exadata Storage Server Software Innovations

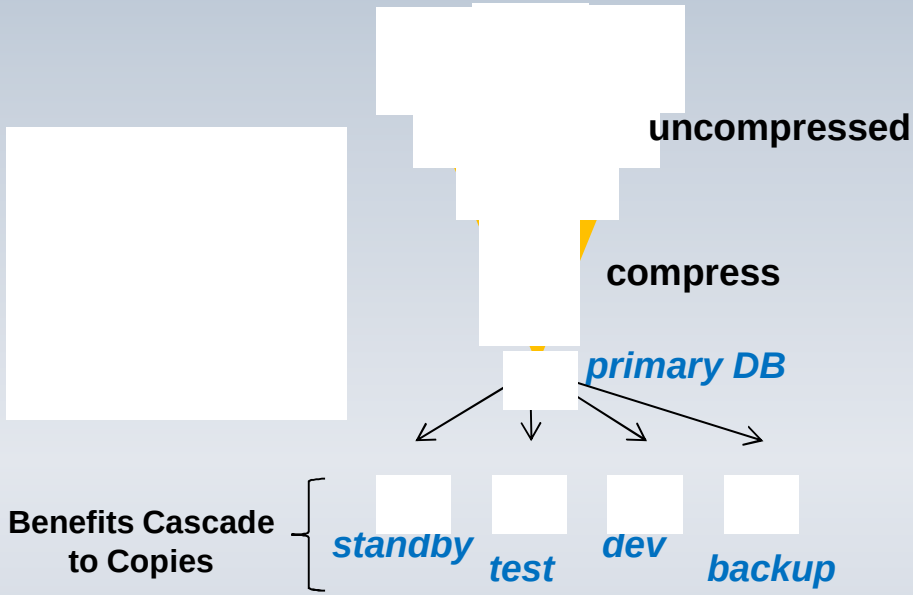
- **Intelligent storage**
  - Scale-out InfiniBand storage
  - Smart Scan query offload



- **Smart PCI Flash Cache**
  - Accelerates random I/O up to 30x
  - Triples data sc



- **Hybrid Columnar Compression**
  - 6-10x compression for warehouses
  - 10-15x compression for archives



Roles for Data Warehouse & Middle Tier BI

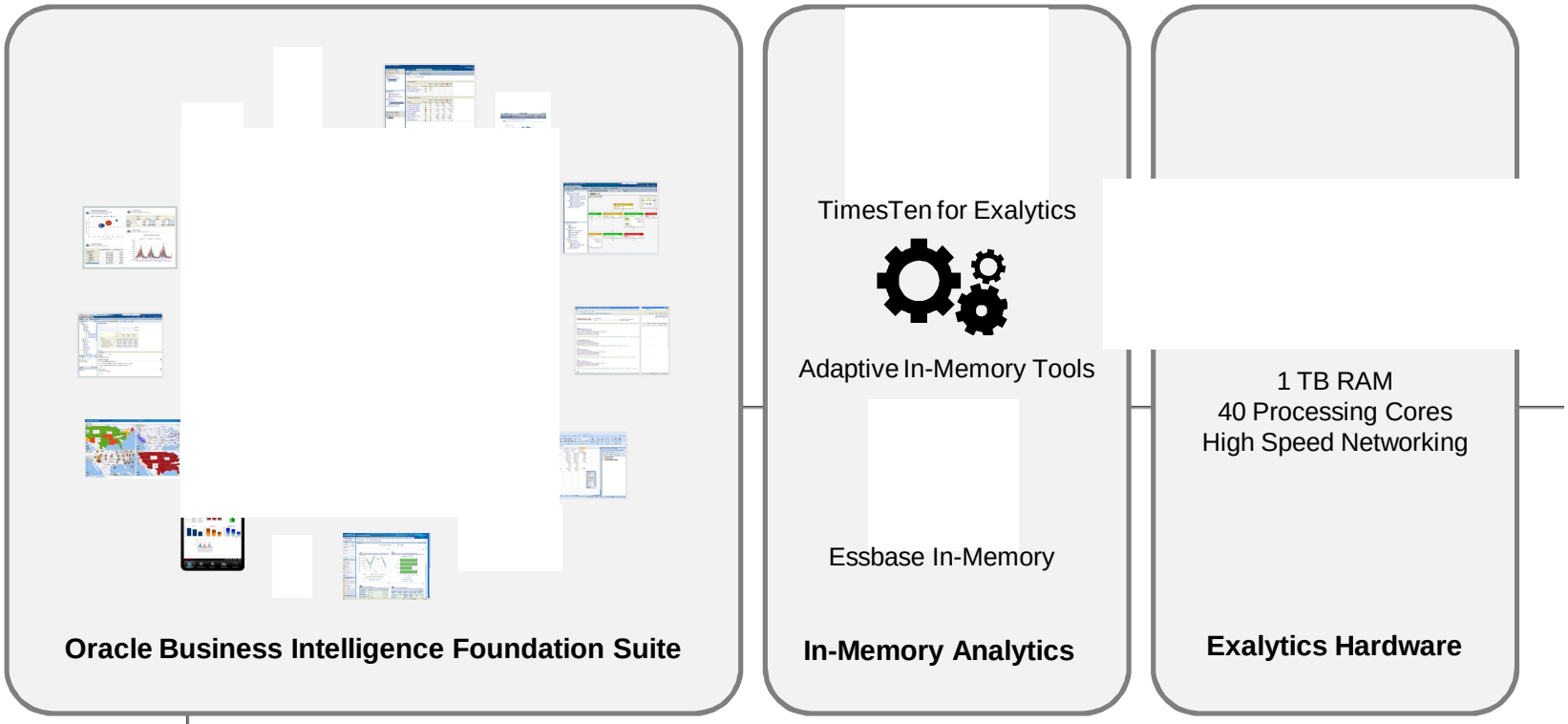
Data Warehouse

- Optimized storage for enterprise data volumes
- Exceeds performance & availability SLAs
- Persistent & secure version of the truth
- Flexible schema
- IT timeframes for solutions

Middle-Tier BI

- Optimized for information delivery
- Quality of data visualization is key
- Discovery, scenario modeling, scorecards
- Dimensional-style self-guided analysis
- Easy to add new sources of data

Oracle Exalytics & BI Foundation Suite Platform



End-user Experience with Exalytics

Speed of Thought Interactive Analysis

- Highly Interactive Analysis
- Free Form Data Exploration
- View Auto Suggestions
- Contextual Actions

TimesTen In-Memory Database for Exalytics

Better Analytics Support

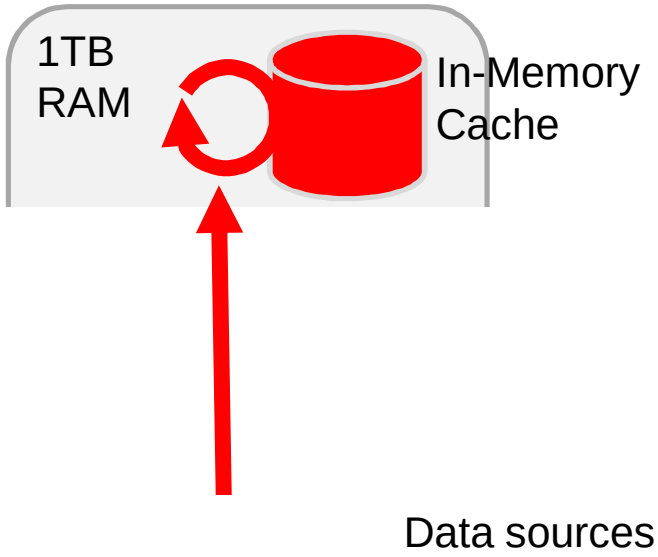
- OLAP Grouping Operators: CUBE, ROLLUP, GROUPING SETS
- WITH Clause
- Analytic Functions: RANK, DENSE_RANK, SUM, AVG, ORDER BY NULLS FIRST|LAST
- Time functions: TIMESTAMPADD, TIMESTAMPDIFF
- Columnar Compression

In-Memory Analytics: Intelligent Cache

Full Navigation Into Result Cache

- Automatically Cache Past Results
- Treat Cache as Logical Table Source
- Re-write queries on the cache
- Applicable to any size DW
- Tools
  - BI Server result cache
  - In-Memory store

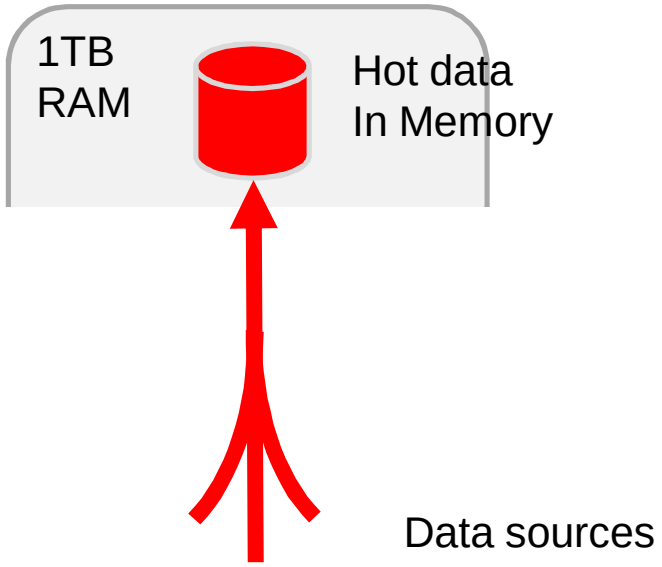
- Install to create cache directories in memory



In-Memory Analytics: Adaptive Data Mart

"Hot" data mart in memory

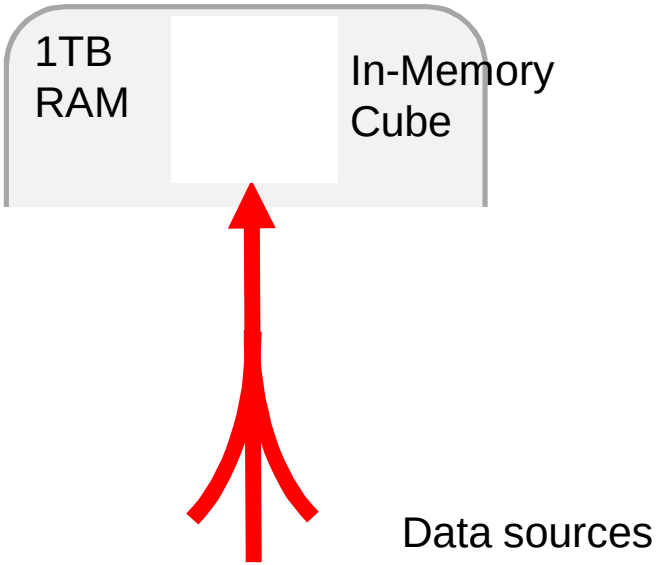
- Automatically identify
  - Slow data sources, facts, grains
  - Workload distribution
  - Optimal data mart for overall performance
- Applicable to any size DW
- Clustering to expand memory
- Tools
  - Summary advisor
  - BI Server aggr. persist. to create 'hot' data mart in TimesTen for Exalytics
  - Incremental refresh: double buffering



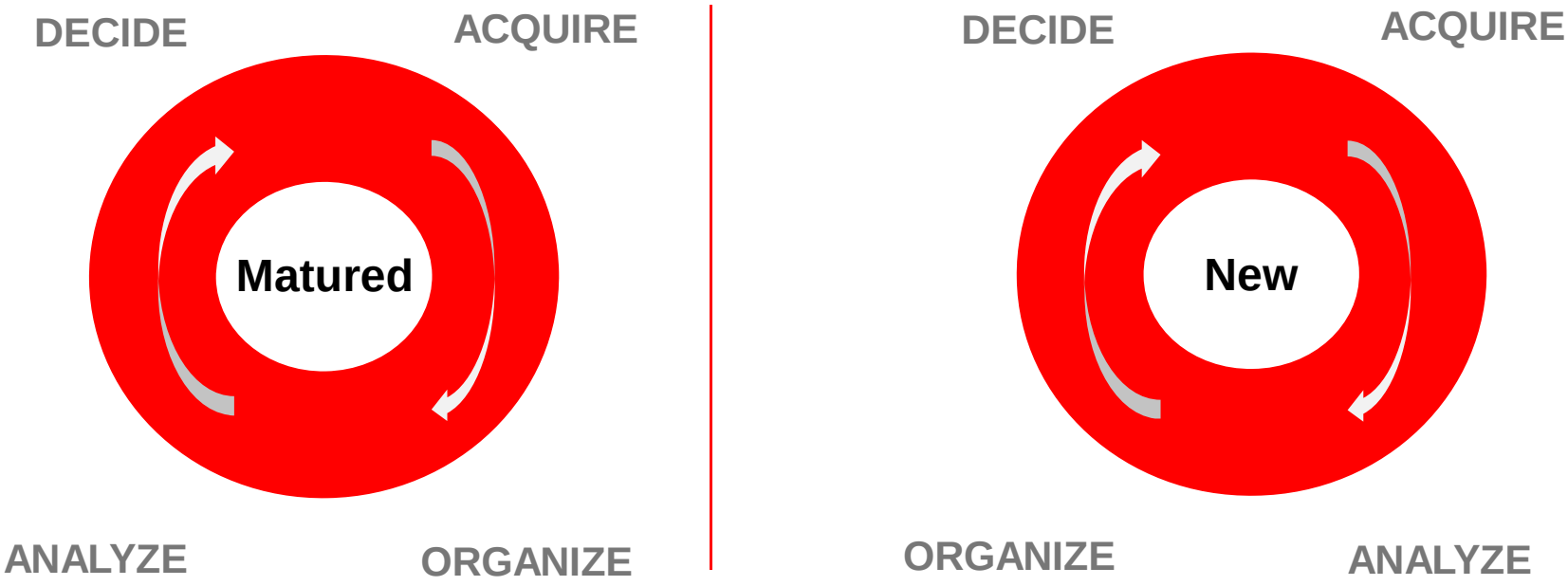
In-Memory Analytics: Essbase Cubes

Specific Subject Areas in Memory

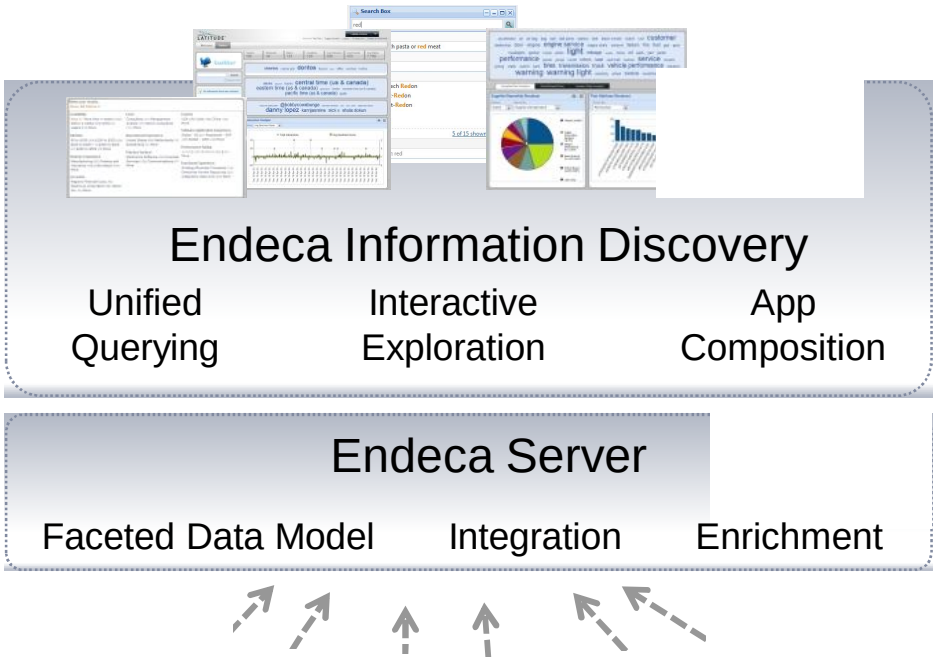
- Specify subject areas for cube spinoff
- High performance scenario modeling (read+write)
- Query acceleration
- Manual configuration



Matured vs. New Data Analysis Processes



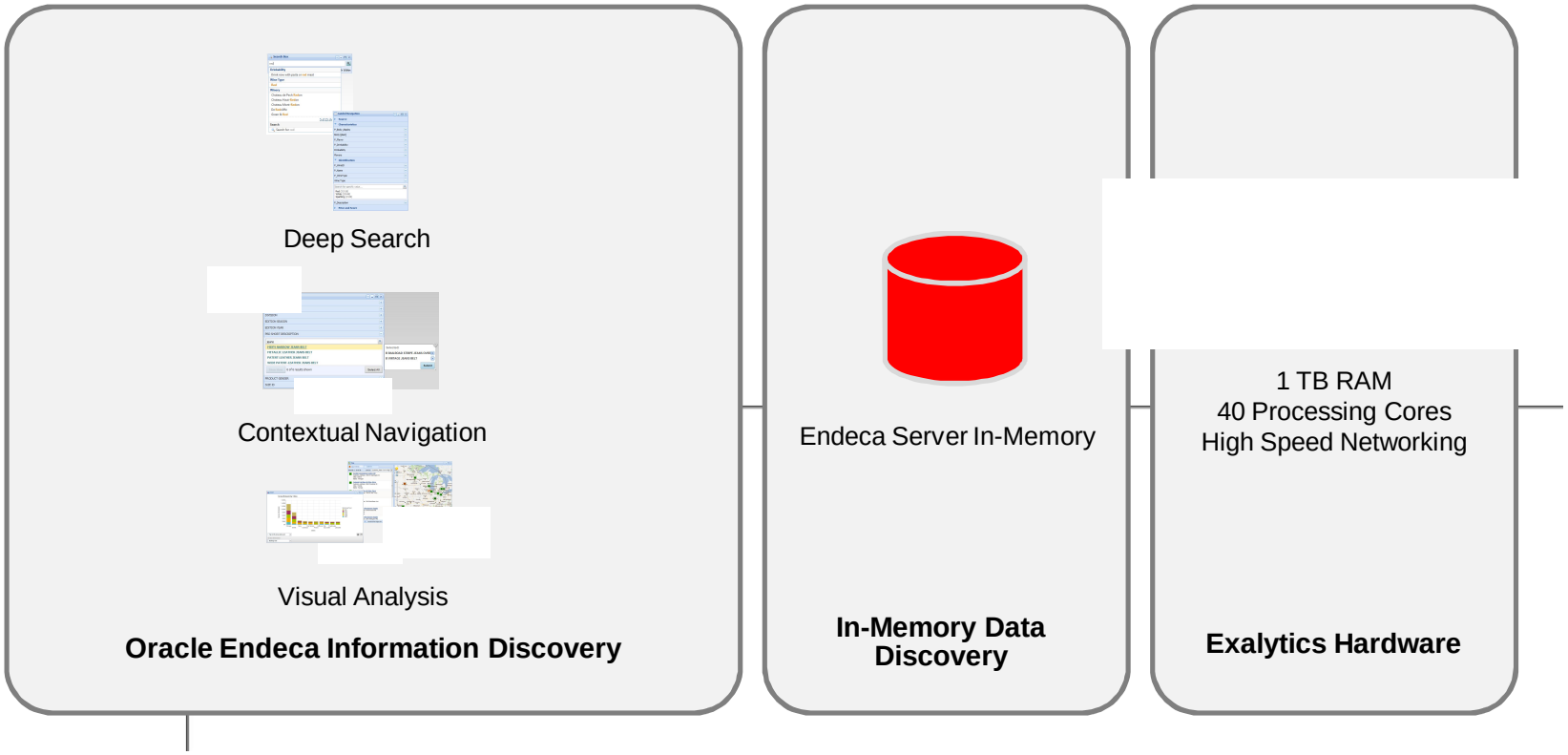
Oracle Endeca Information Discovery



Helps organizations quickly explore **ALL** relevant data

- Combines structured & unstructured data from disparate systems
- Automatically organizes information for search, discovery & analysis

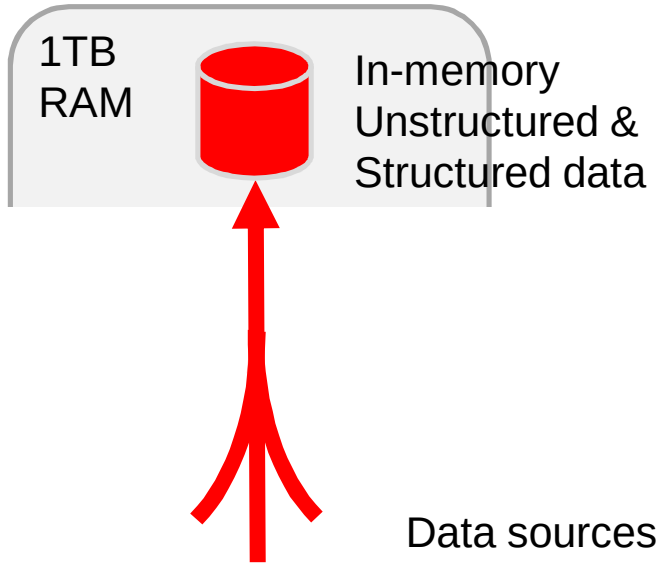
Oracle Exalytics & Endeca Platform



In-Memory Data Discovery: Endeca Server

Unstructured & Structured Data in Memory

- Specify data sources for loading into Endeca Server
- High performance data discovery for all types of data in the Endeca Server
- Manual configuration



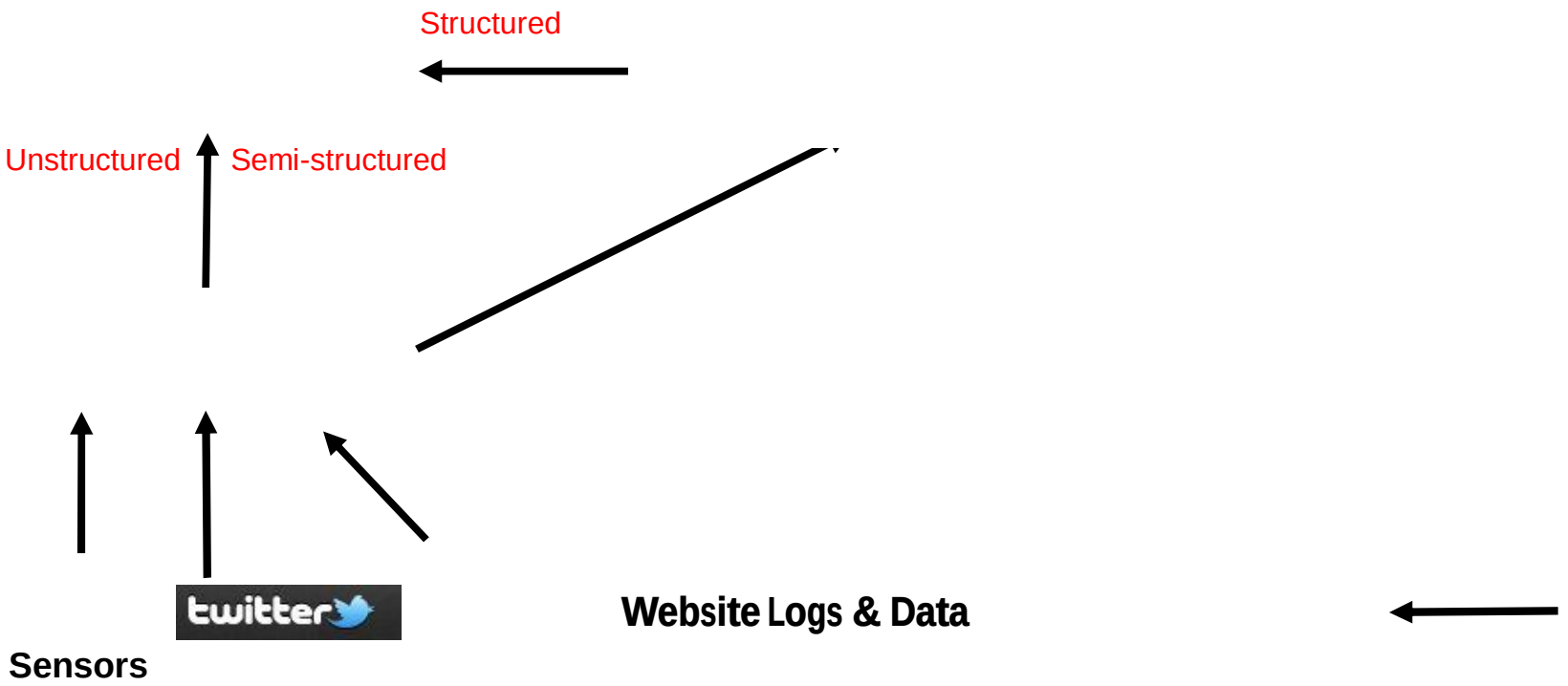
Making Sense of Diverse Data Sources

Shopping
Cart Site

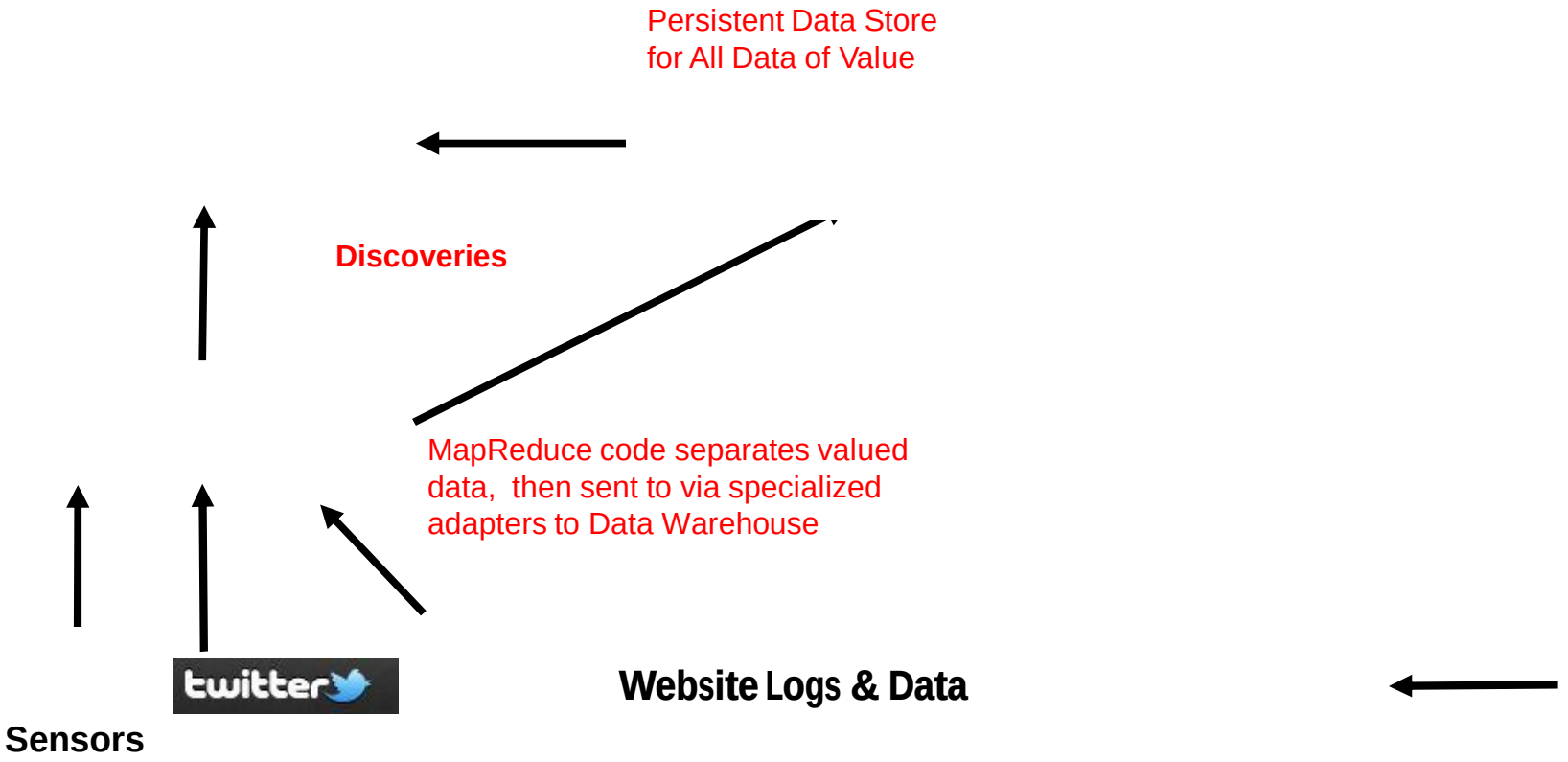
Website Logs & Data

Sensors

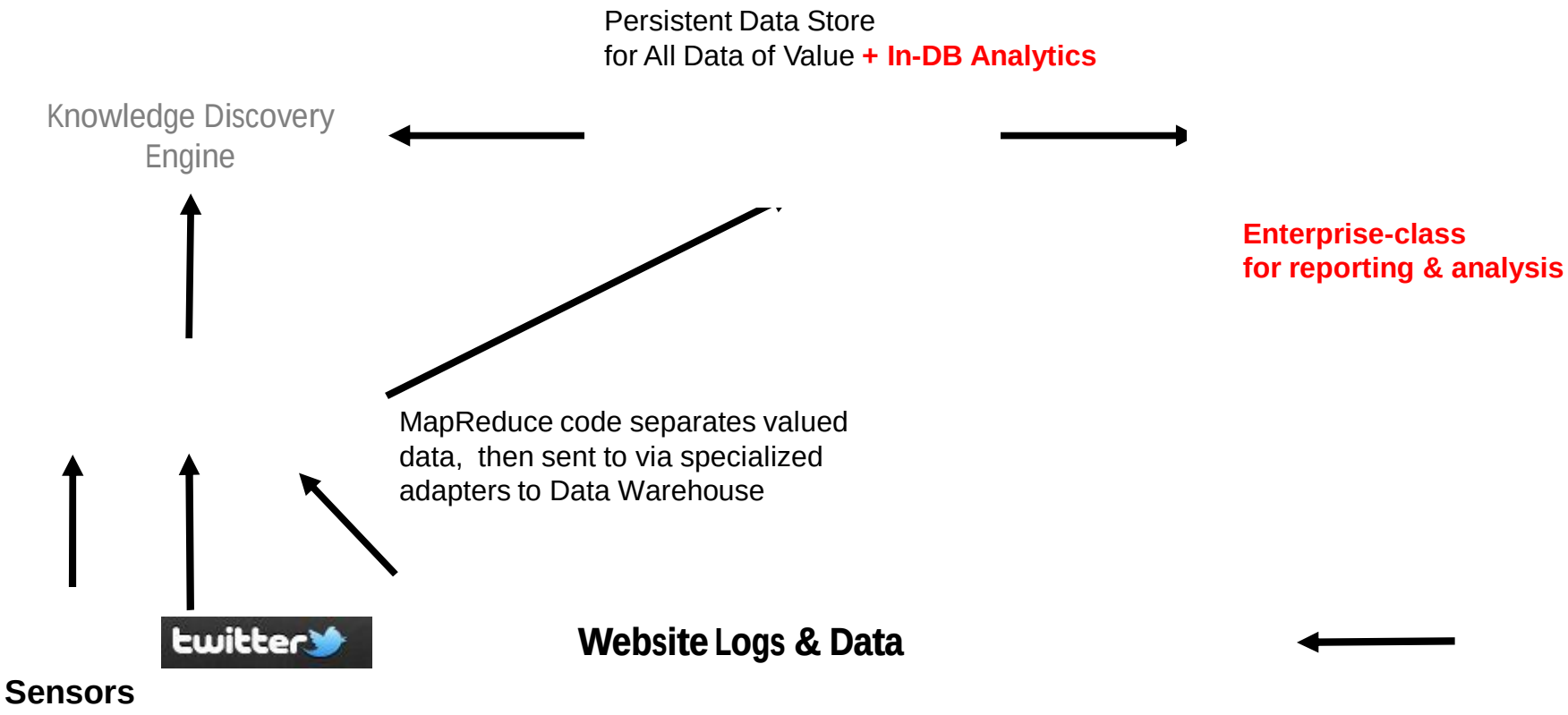
Determine Value of Data of All Types



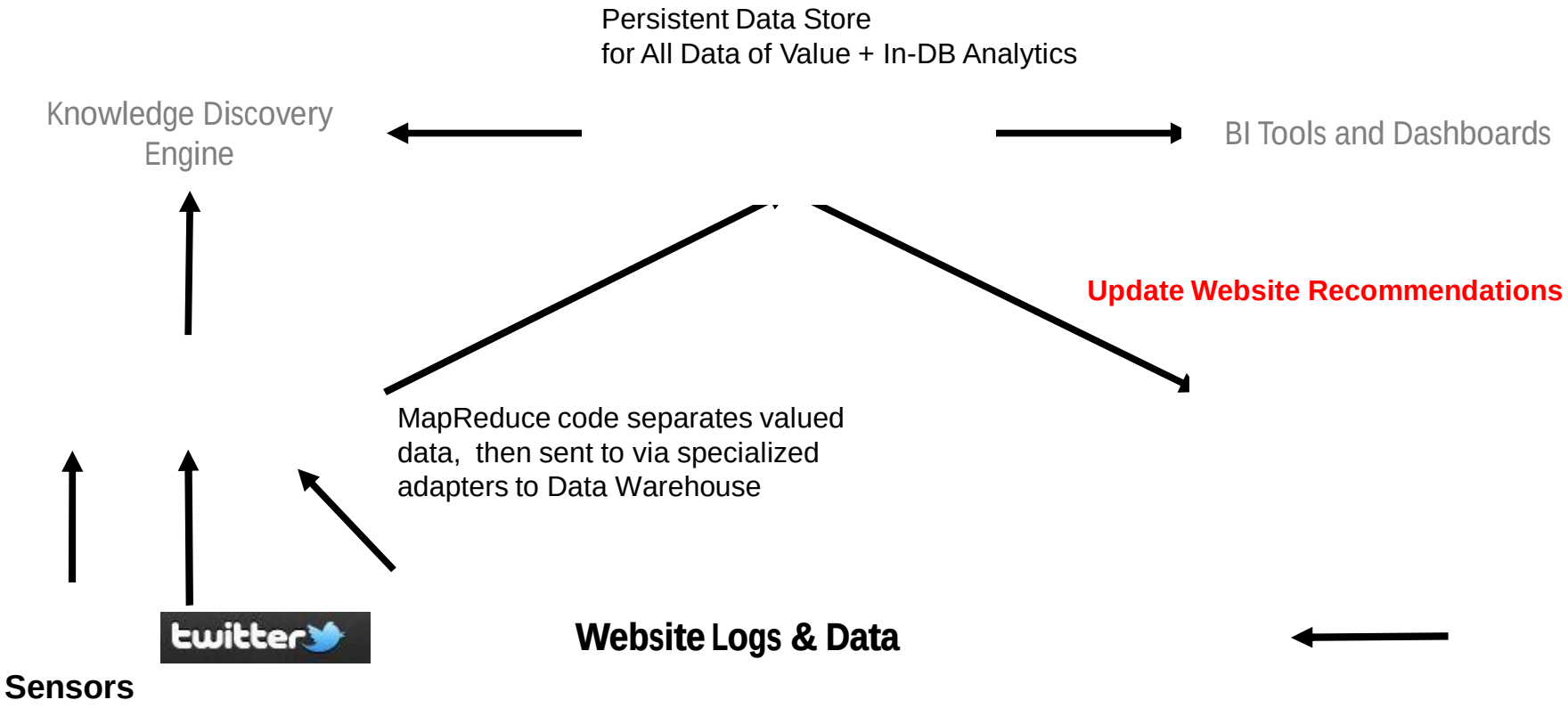
Valuable Data Found – Now Store it Securely



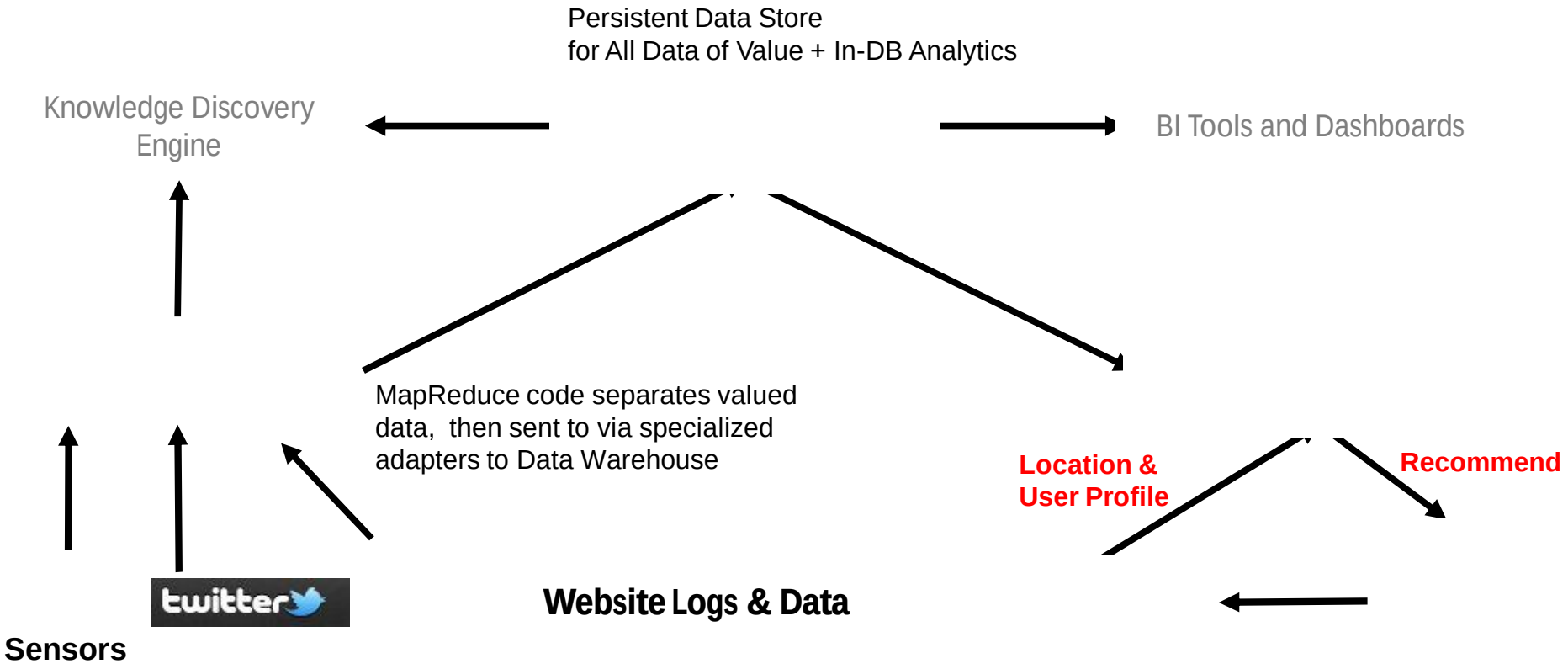
Deploy Widely Available Reports & Analytics



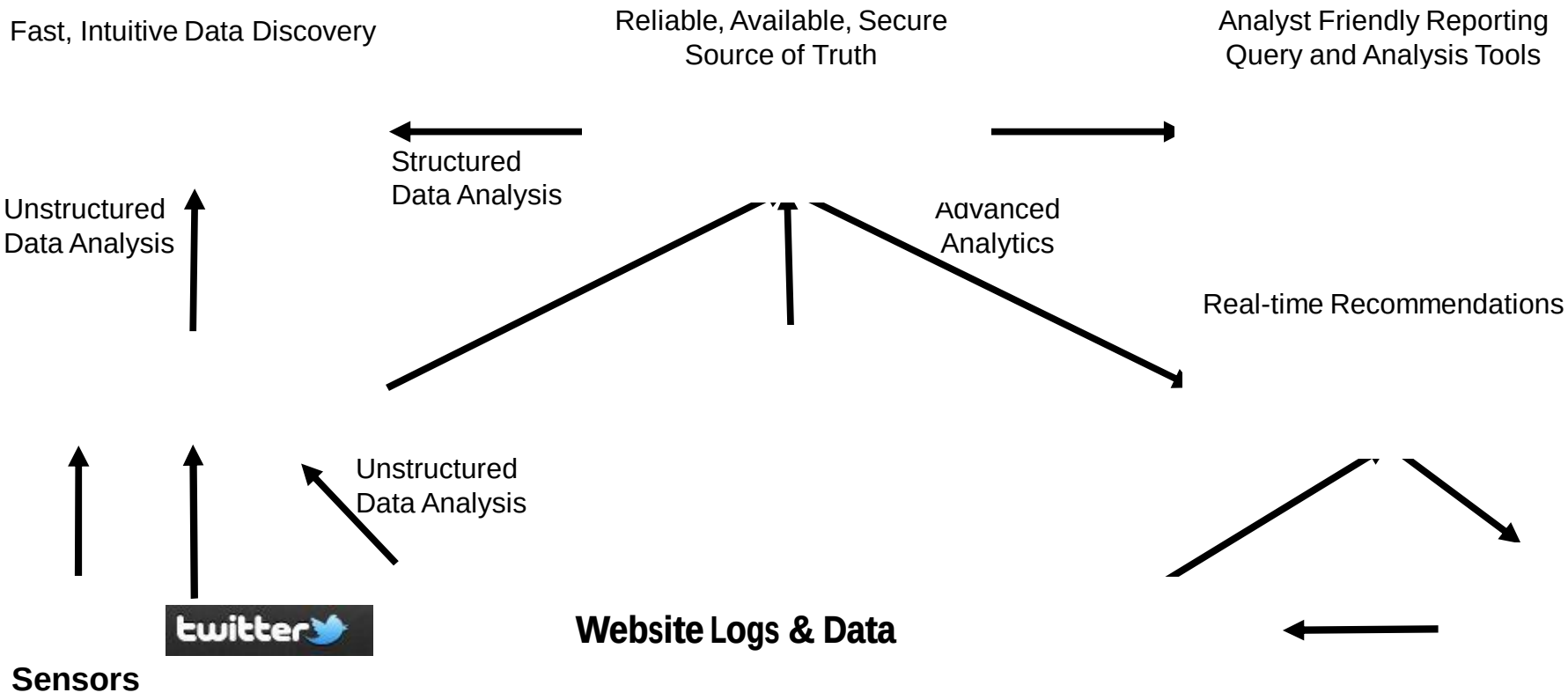
Feed the Recommendation Engine



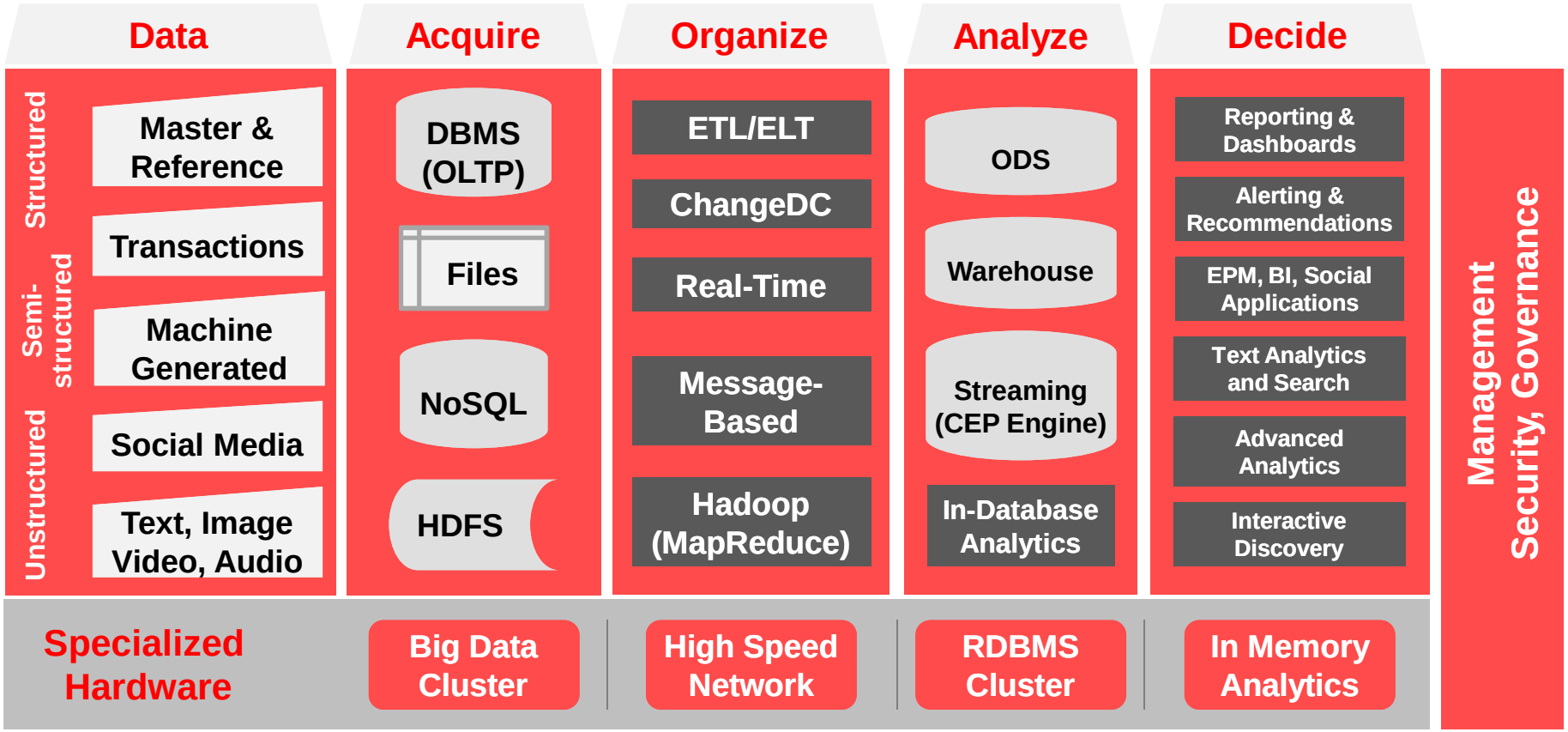
Make Well-Tuned Real-Time Recommendations



Only Oracle Offers this Entire Solution



Oracle Big Data Architecture Capabilities



Oracle Delivers Value from ALL Data

Best for Business, Best for IT

Better Insights, Decisions, Actions

- From measurement to analysis, forecasting & optimization
- Insights across time, functions and roles
- Persistent version of the truth for **ALL DATA**

Most Complete, Open, Integrated

- From Discovery to Dashboards to Analytics to Data Management
- Standards based & blending of Open Source components
- Optimized integrated Engineered Systems & software

World Class Analytics Infrastructure

- Best of Breed capabilities at each layer of the stack
- Uniquely enables complete analysis of **ALL DATA**
- Enterprise Architecture: scalable, reliable, manageable & secure

Q&A