

ORACLE®

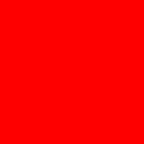


ORACLE[®]

Oracle's Big Data solutions

Jean-Philippe Breysse

Oracle Suisse



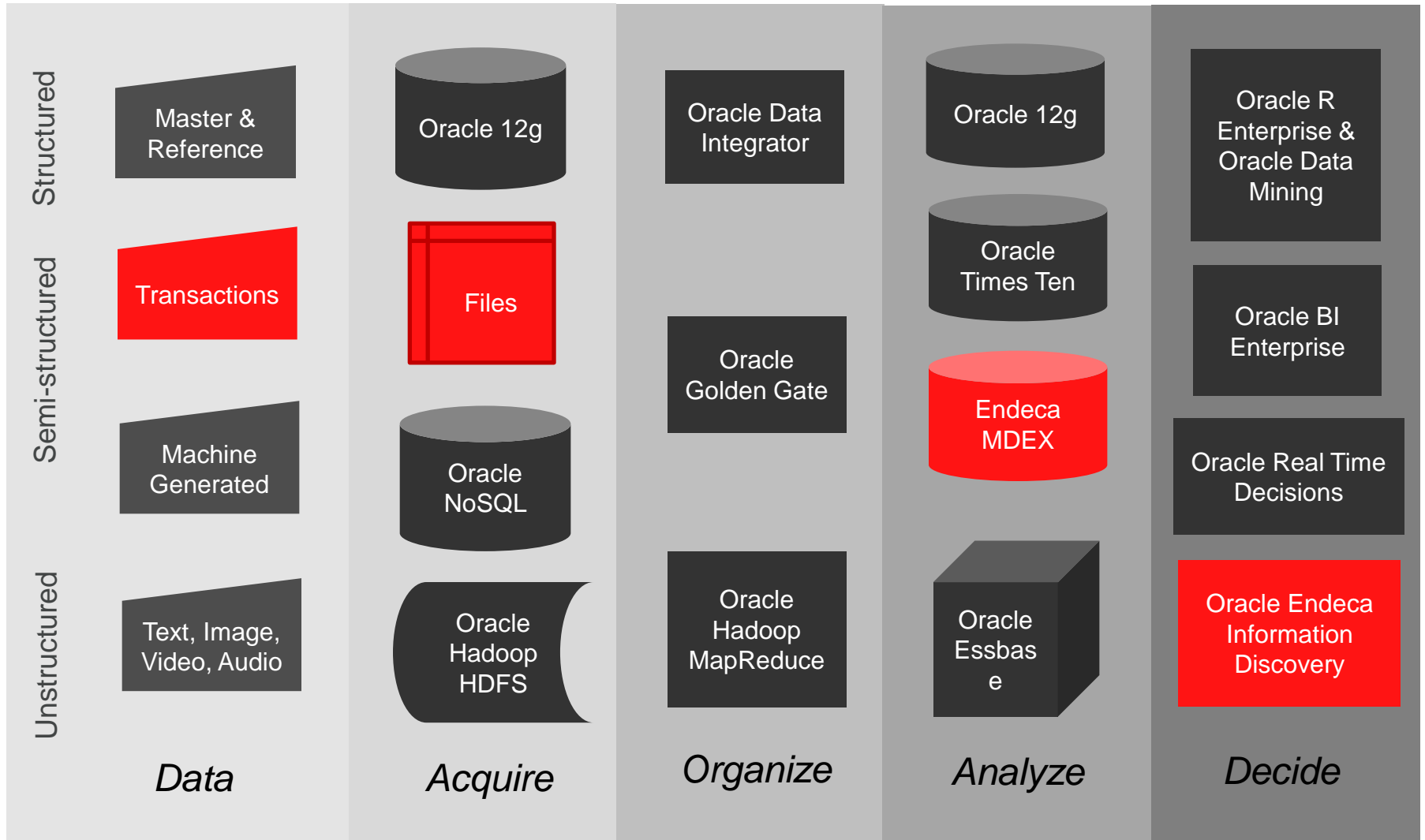
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.

The development, release, and timing of any features or functionality described for Oracle's products remain at the sole discretion of Oracle.

USE CASE 3: LOGS ANALYSIS OF SERVERS

- Short Description :
 - Daily logs analysis
- Issues:
 - Find correlations on what drives to failures
 - Log files stored as flat files

Oracle Technology mapped to Analytics Landscape



Agenda

- Big Data
- Solution Spectrum
- Inside the Big Data Appliance
- Big Data Applications Software
- Big Data Analytics
- Conclusions

Big Data

Why Everyone Should Care



Tapping into Diverse Data Sets

Big Data:
Decisions based
on all your data

Video and Images



Documents



Social Data



Machine-Generated Data



**Information
Architectures
Today:**

Decisions based
on database data


Transactions

ORACLE
FUSION APPLICATIONS

ORACLE
SIEBEL

SAP

A bit of history ...

 **hadoop**: Developed initially by Doug Cutting (*Nutch - Open-source websearch engine*) and Yahoo -> inspired by Google's papers on MapReduce and GFS (2003-2004) resulted in Apache Hadoop (2006)

Amazon Dynamo (2007): distributed systems technologies

Cassandra: was developed at Facebook (2008) to power their Inbox Search feature (columnar oriented distributed DB) based initially on Dynamo and Bigtable (built by Google)

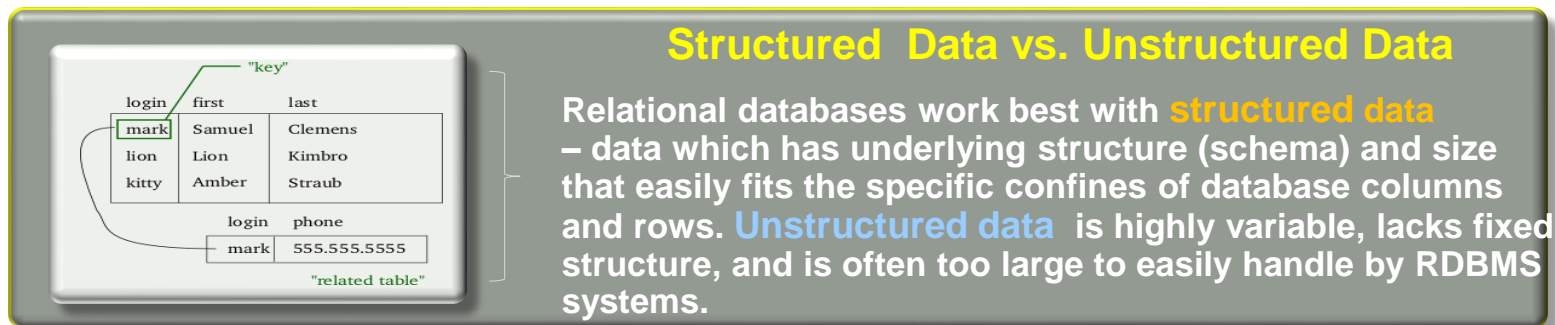
Voldemort: is a distributed data store that is designed as a key-value store used by LinkedIn for high-scalability storage (NoSql key value)

Cloudera: . It contributes to Hadoop and related Apache projects and provides a commercial distribution of Hadoop

So What is Big Data Anyway?

It's a matter of perspective. Big Data is both:

- **LARGE AND VARIABLE DATASETS** that are difficult for traditional database tools to easily manage – including datasets that once seemed not important or too problematic to deal with. Big Data datasets include:
 - Extremely large files of *unstructured or semi-structured* data
 - Large and highly distributed datasets that are otherwise difficult to manage as a single unit of information
- **NEW SET OF TECHNOLOGIES** that can economically capture, store, manage, and extract value from Big Data datasets – thus facilitating better, more informed business decisions



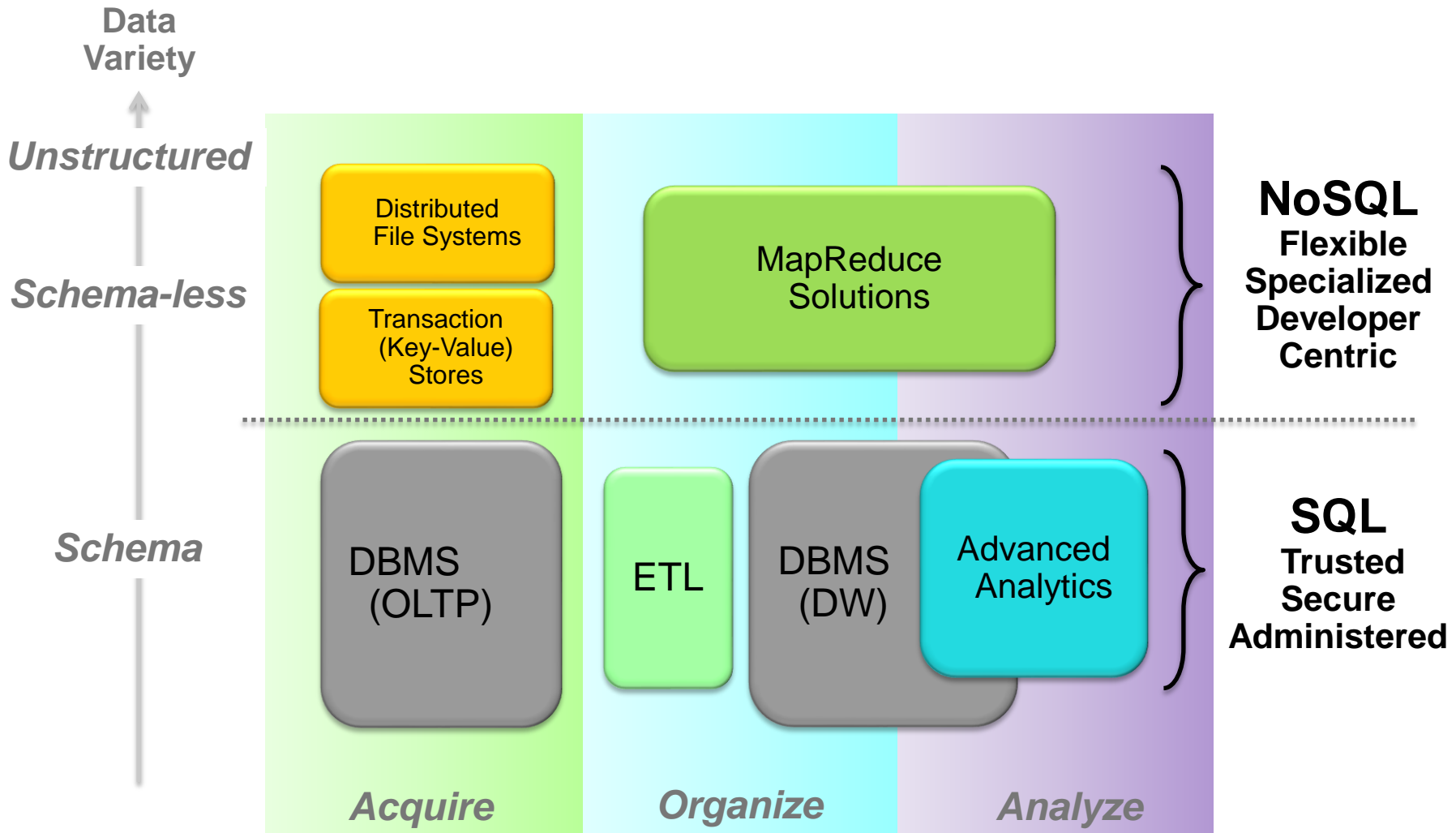
Source: IDC Digital Universe Study, Extracting Value from Chaos, June 2011 (sponsored by EMC)

Drive Value from Big Data

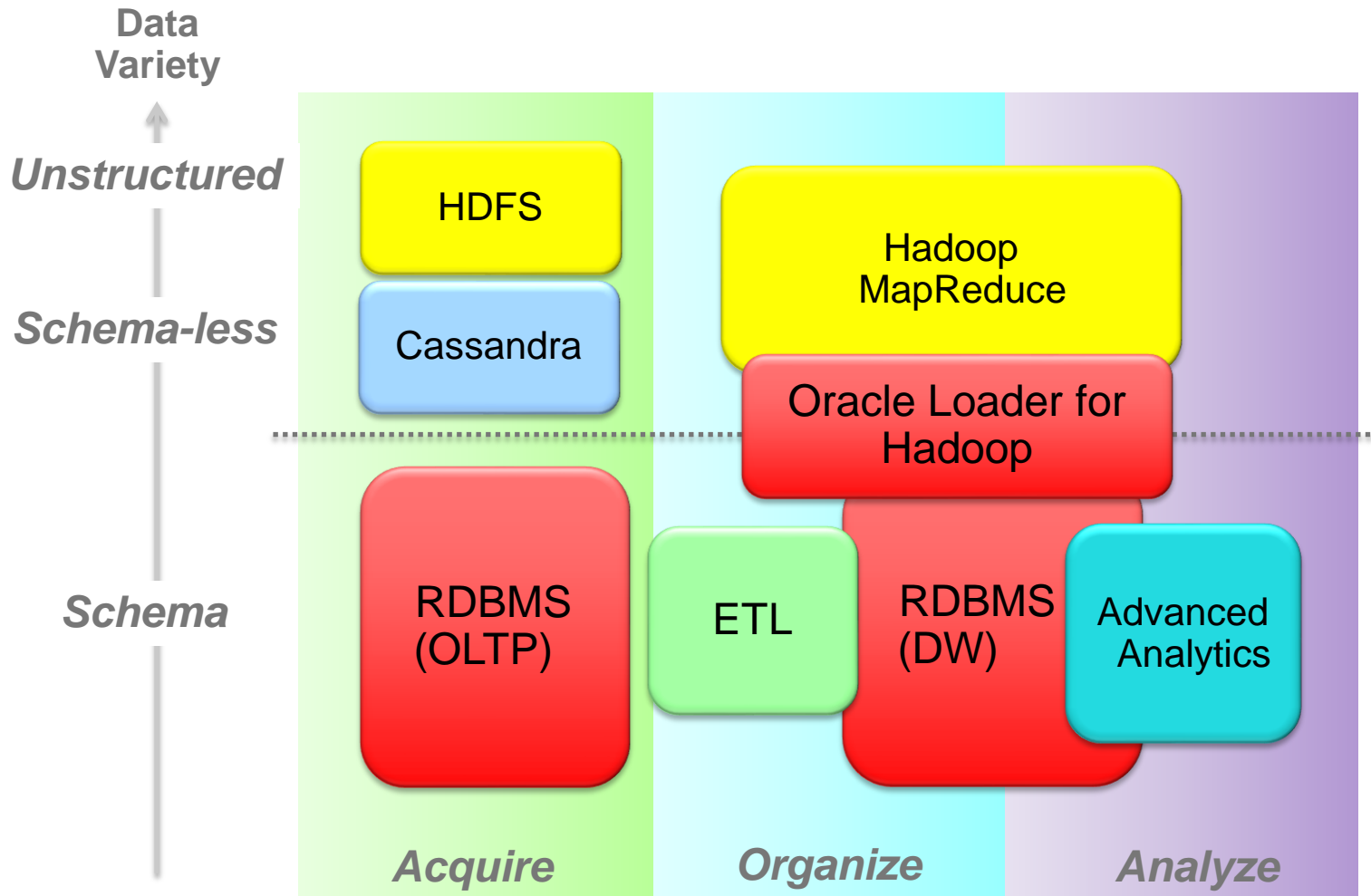
Building a Big Data Platform



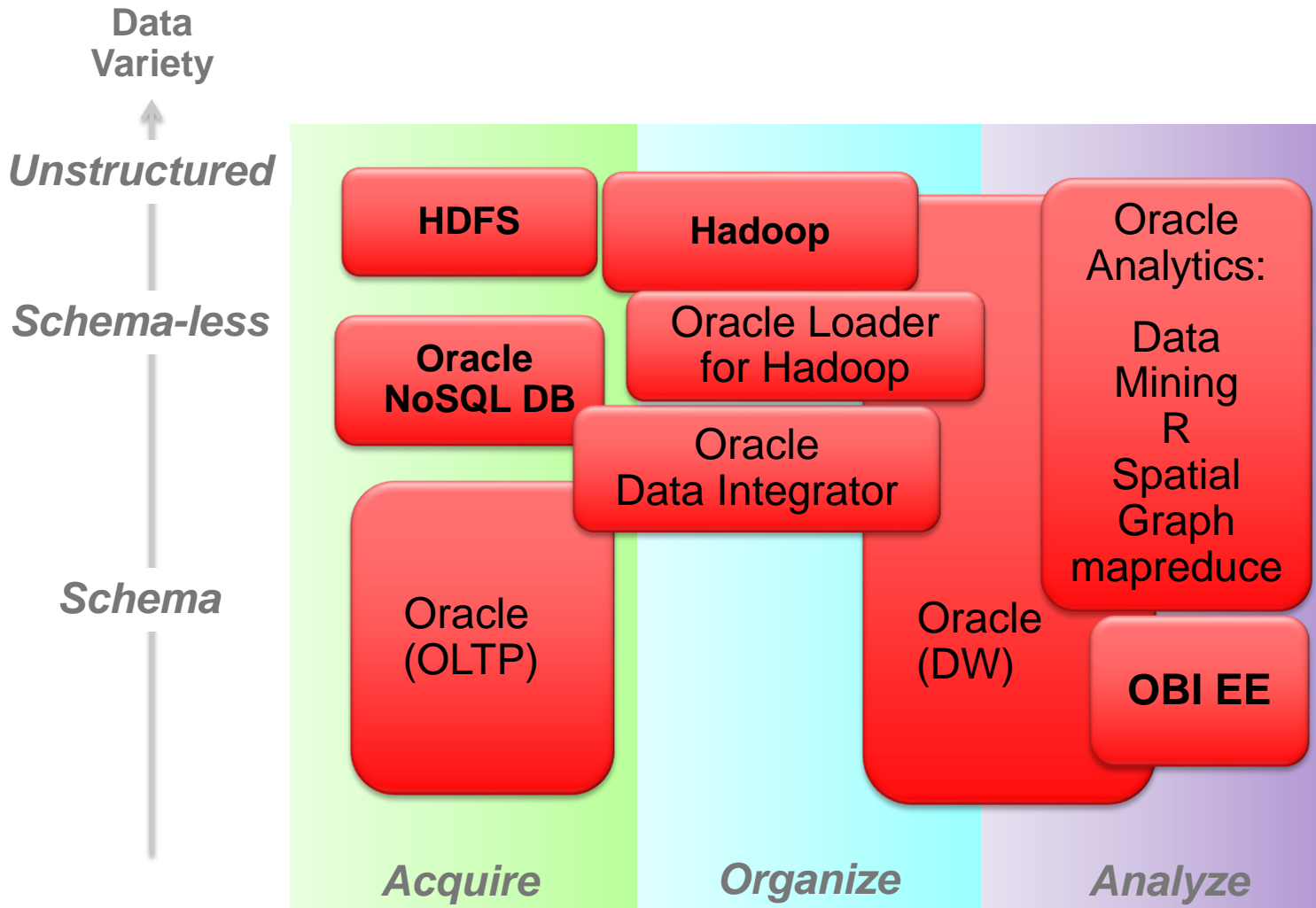
Divided Solution Spectrum



Hadoop to Oracle – Bridging the Gap



Oracle Integrated Software Solution

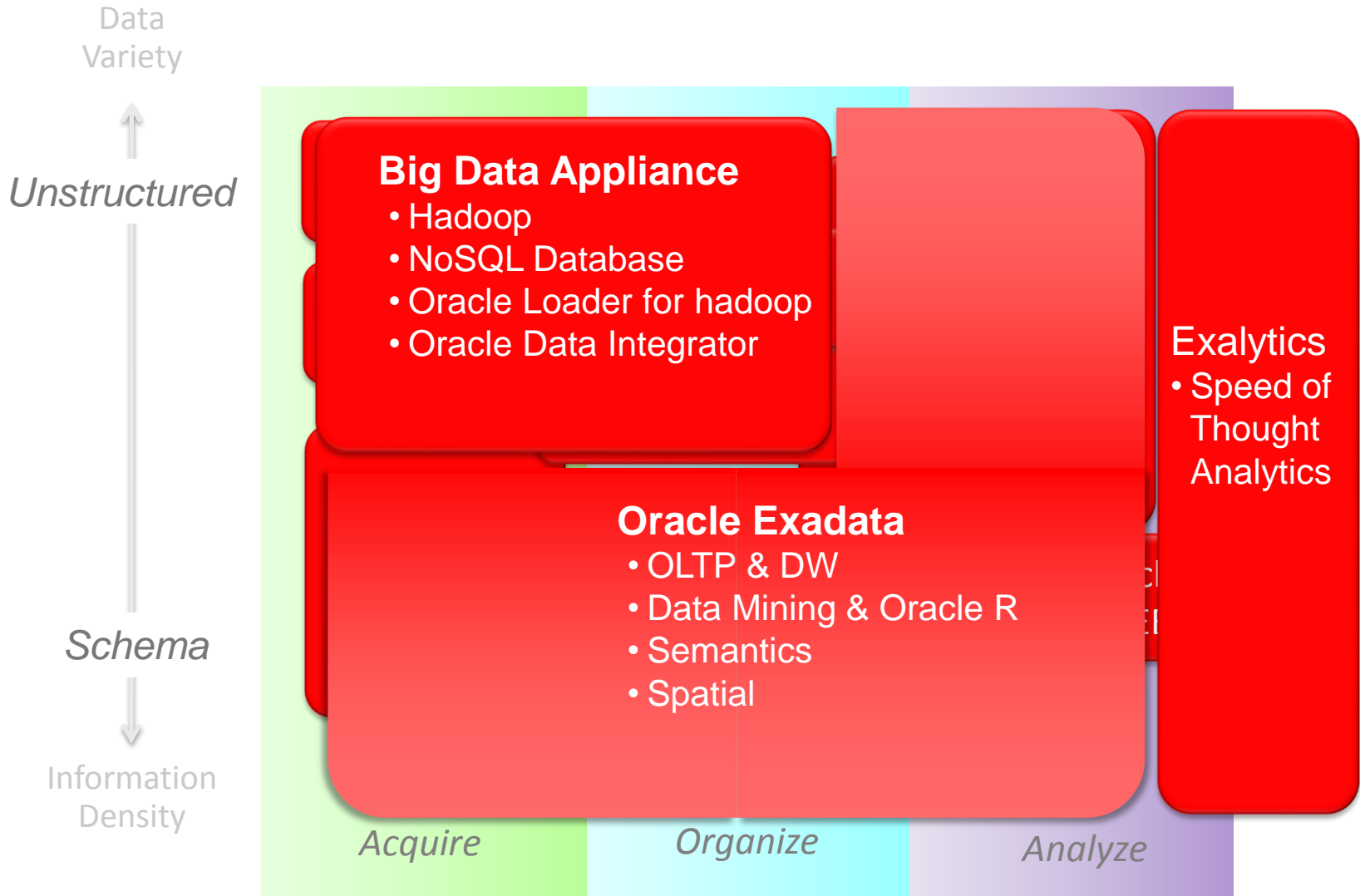


Inside the Big Data Appliance

Overview

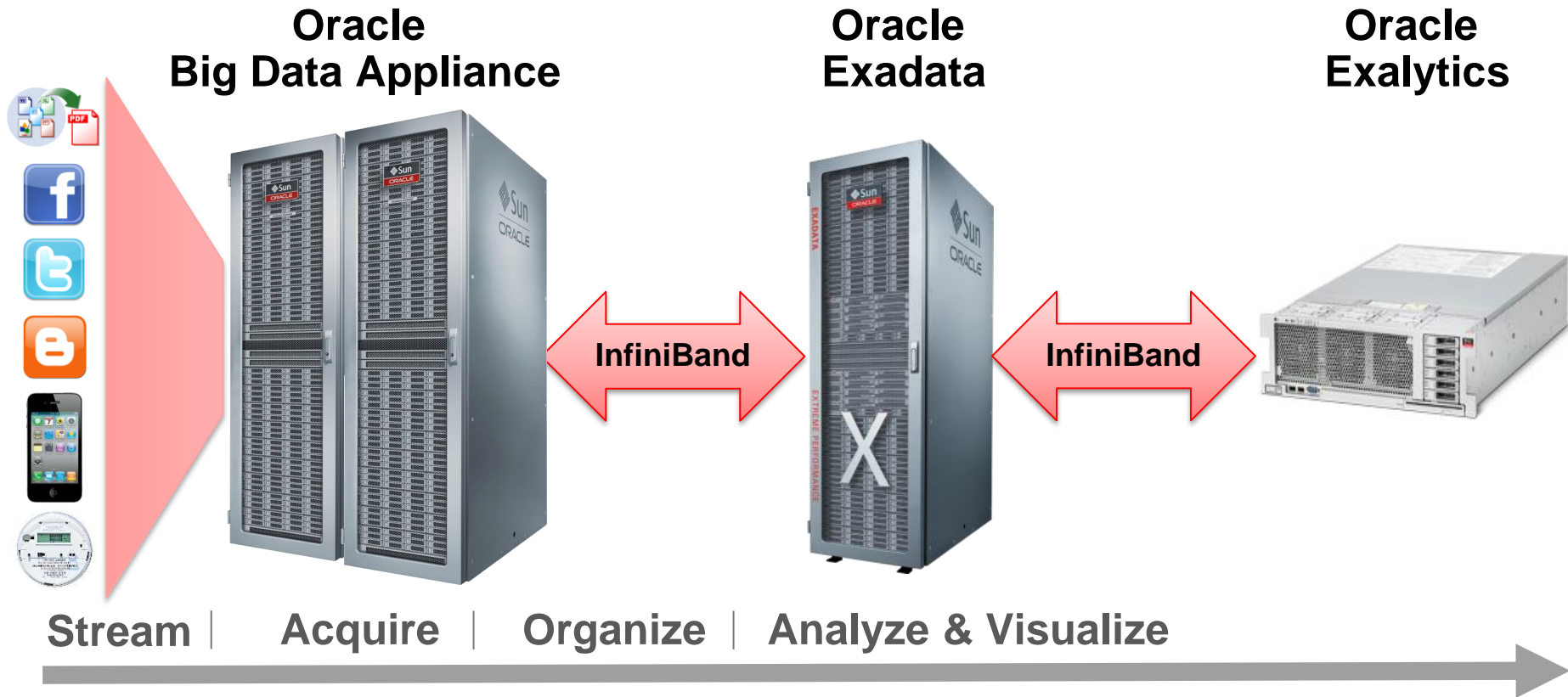


Oracle Engineered Solutions



Big Data Appliance

Usage Model



Oracle Big Data Appliance Hardware

- 18 Sun X4270 M2 Servers
 - 48 GB memory per node = 864 GB memory
 - 12 Intel cores per node = 216 cores
 - 24 TB storage per node = 432 TB storage
- 40 Gb p/sec InfiniBand
- 10 Gb p/sec Ethernet



Scale Out to Infinity



Scale out by connecting racks to each other using Infiniband

- Expand up to eight racks without additional switches
- Scale beyond eight racks by adding an additional switch

Oracle Big Data Appliance Software

- Oracle Enterprise Linux 5.6
- Oracle Hotspot Java VM
- Cloudera's Distribution including Apache Hadoop
- Cloudera Manager
- Open Source Distribution of R
- Oracle NoSQL Database Community Edition



Big Data Application Software

Acquire New Information



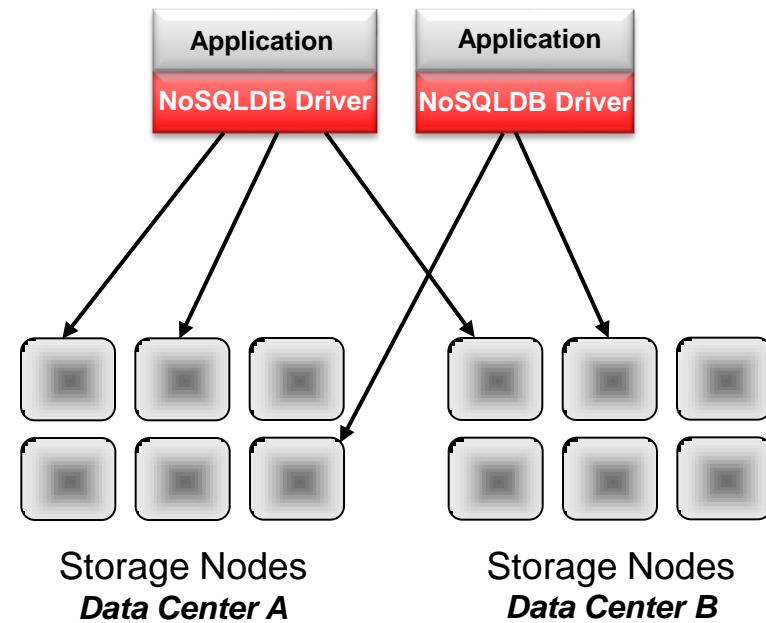
Key-Value Store Workloads

- Large dynamic schema based data repositories
- Data capture
 - Web applications (click-through capture)
 - Online retail
 - Sensor/statistics/network capture (factory automation for example)
 - Backup services for mobile devices
- Data services
 - Scalable authentication
 - Real-time communication (MMS, SMS, routing)
 - Personalization
 - Social Networks

Oracle NoSQL DB

A distributed, scalable key-value database

- Simple Data Model
 - Key-value pair with major+sub-key paradigm
 - Read/insert/update/delete operations
- Scalability
 - Dynamic data partitioning and distribution
 - Optimized data access via intelligent driver
- High availability
 - One or more replicas
 - Disaster recovery through location of replicas
 - Resilient to partition master failures
 - No single point of failure
- Transparent load balancing
 - Reads from master or replicas
 - Driver is network topology & latency aware
- Elastic (Planned for Release 2)
 - Online addition/removal of Storage Nodes
 - Automatic data redistribution



Big Data Application Software

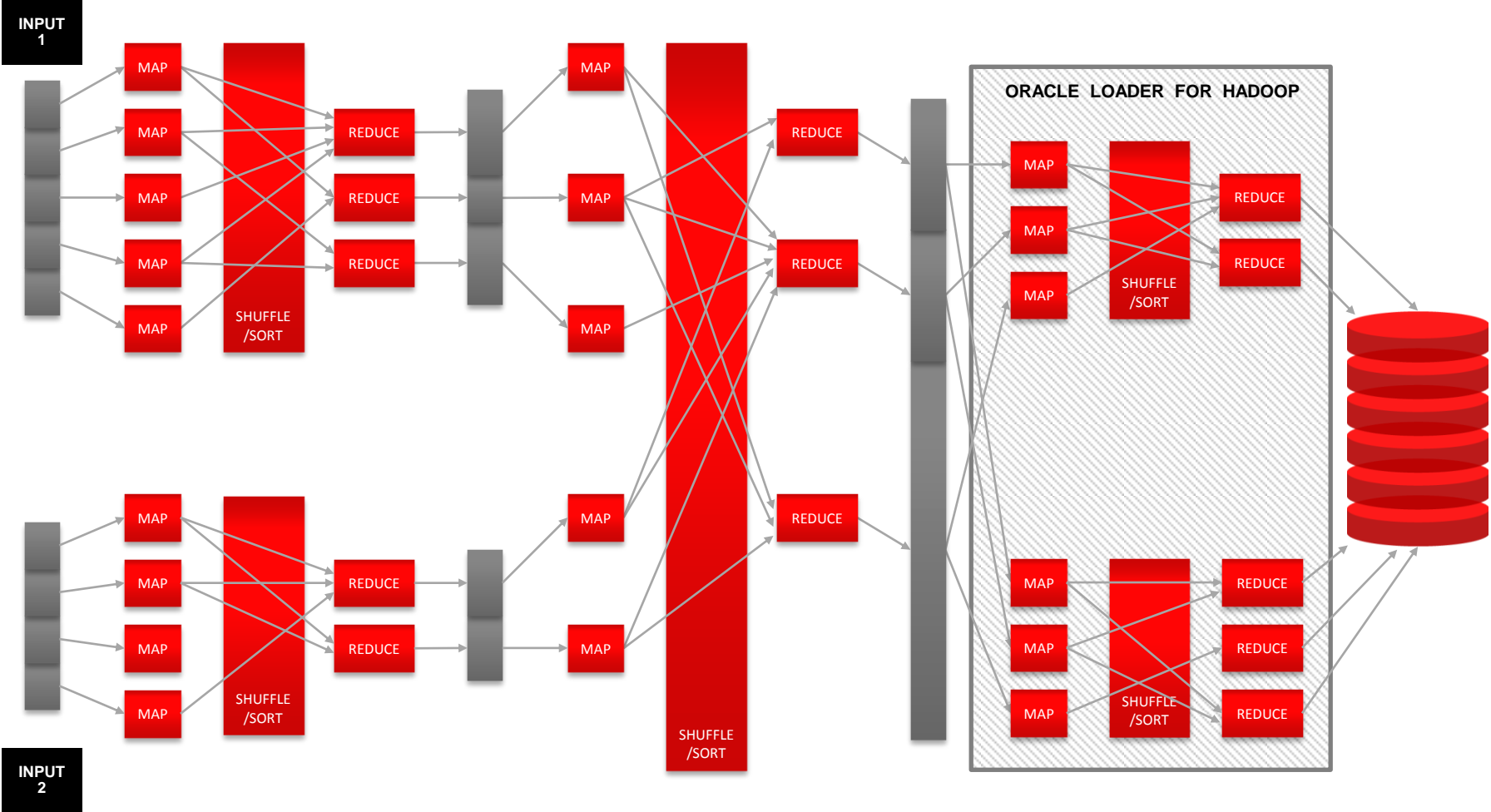
Organizing Data for Analysis



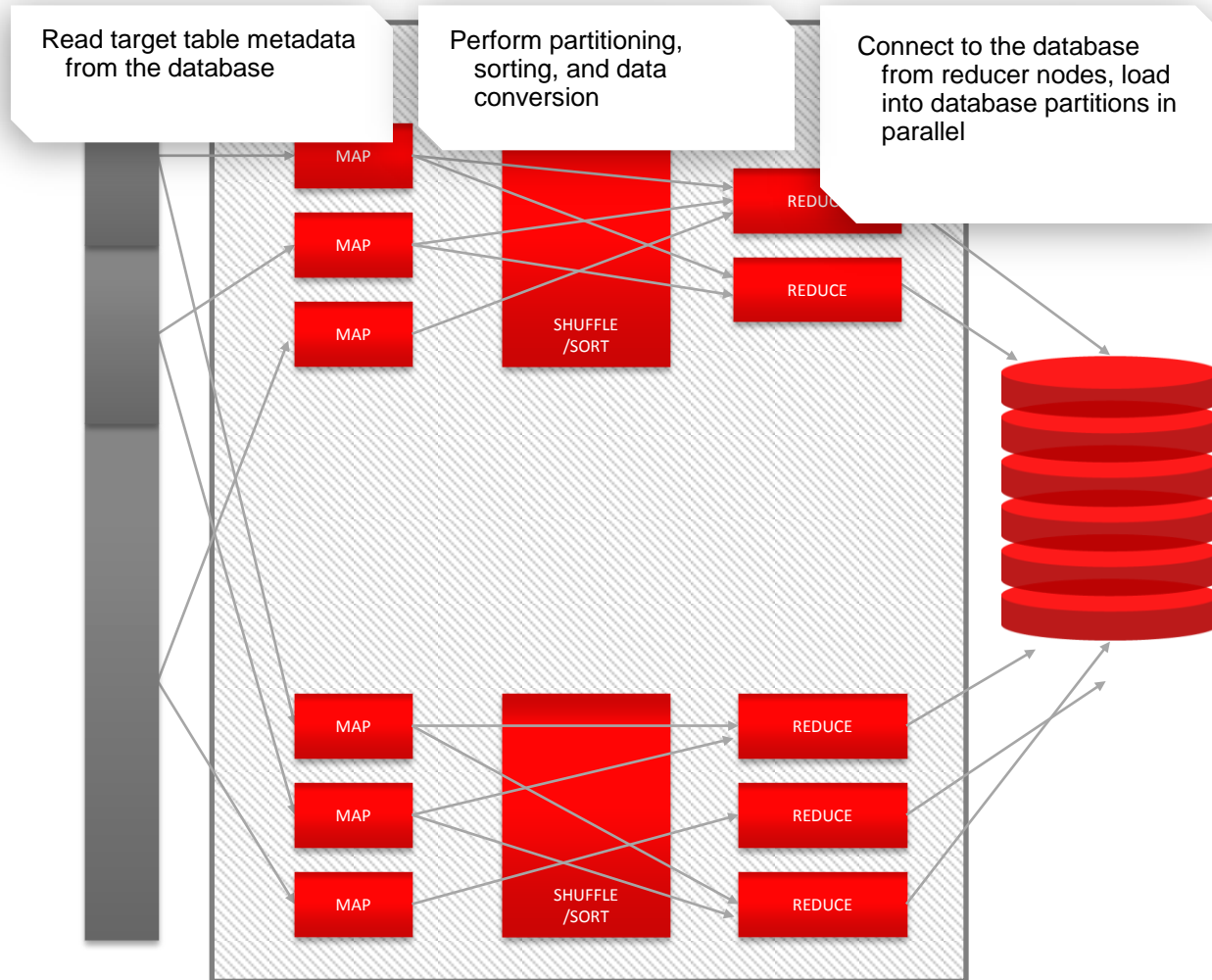
Oracle Loader for Hadoop Features

- Load data into a partitioned or non-partitioned table
 - Single level, composite or interval partitioned table
 - Support for scalar datatypes of Oracle Database
 - Load into Oracle Database 11g Release 2
- Runs as a Hadoop job and supports standard options
- Pre-partitions and sorts data on Hadoop
- Online and offline load modes

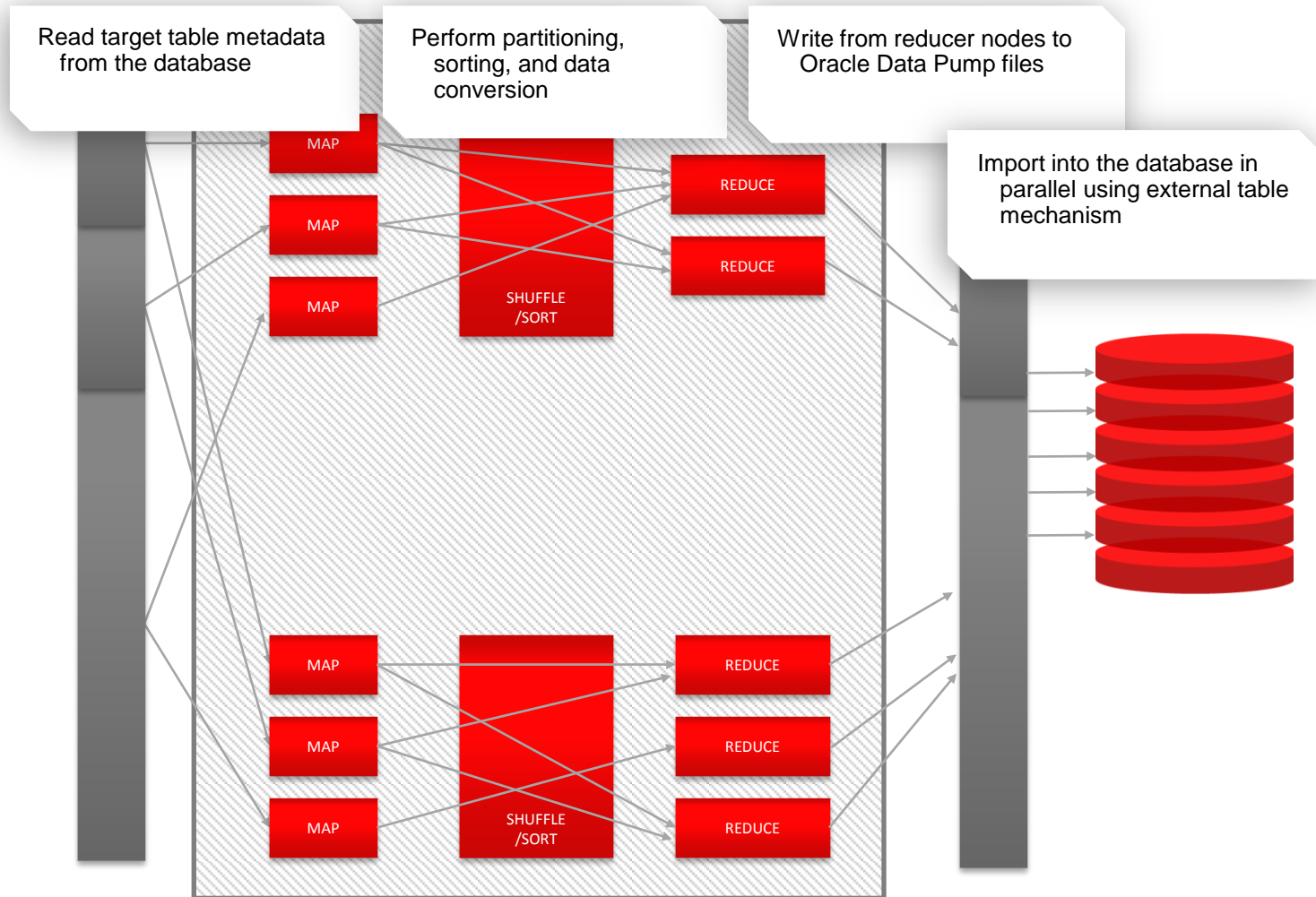
Oracle Loader for Hadoop



Oracle Loader for Hadoop: Online Option



Oracle Loader for Hadoop: Offline Option



Selection Output Option for Use Case

Oracle Loader for Hadoop Output Option	Use Case Characteristics
Online load with JDBC	The simplest use case for non partitioned tables
Online load with Direct Path	Fast online load for partitioned tables
Offline load with datapump files	Fastest load method for external tables
On Oracle Big Data Appliance Direct HDFS	Leave data on HDFS Parallel access from database Import into database when needed

Automate Usage of Oracle Loader for Hadoop

Oracle Data Integrator (ODI)

- ODI has knowledge modules to
 - Generate data transformation code to run on Hive/Hadoop
 - Invoke Oracle Loader for Hadoop
- Use the drag-and-drop interface in ODI to
 - Include invocation of Oracle Loader for Hadoop in any ODI packaged flow

[Smart Mall Data Analysis] Oracle Data Integrator 11g : Process Weblogs

File Edit View Search ODI Tools Window Help

Process Weblogs x 050 Load Oracle x

Designer

Operator

Toolbox

- All
- Changed Data Capture
- Event Detection
- Files
- Internet
- OdiFtpGet
- OdiFtpPut
- OdiInvokeWebService
- Metadata
- Oracle Data Integrator Objects
- Plugins
- SAP
- Utilities

010 Load Weblog → ok → 020 Preprocess → ok → 030 Sessionize → ok → 040 Condense → ok → 050 Load Oracle

Properties

General Advanced Memo Version Privileges

Step name	050 Load Oracle	Type	Flow (Interface)
Linked object	050 Load Oracle	Path	[Smart Mall Data Analysis].[Demo]

Overview Diagram

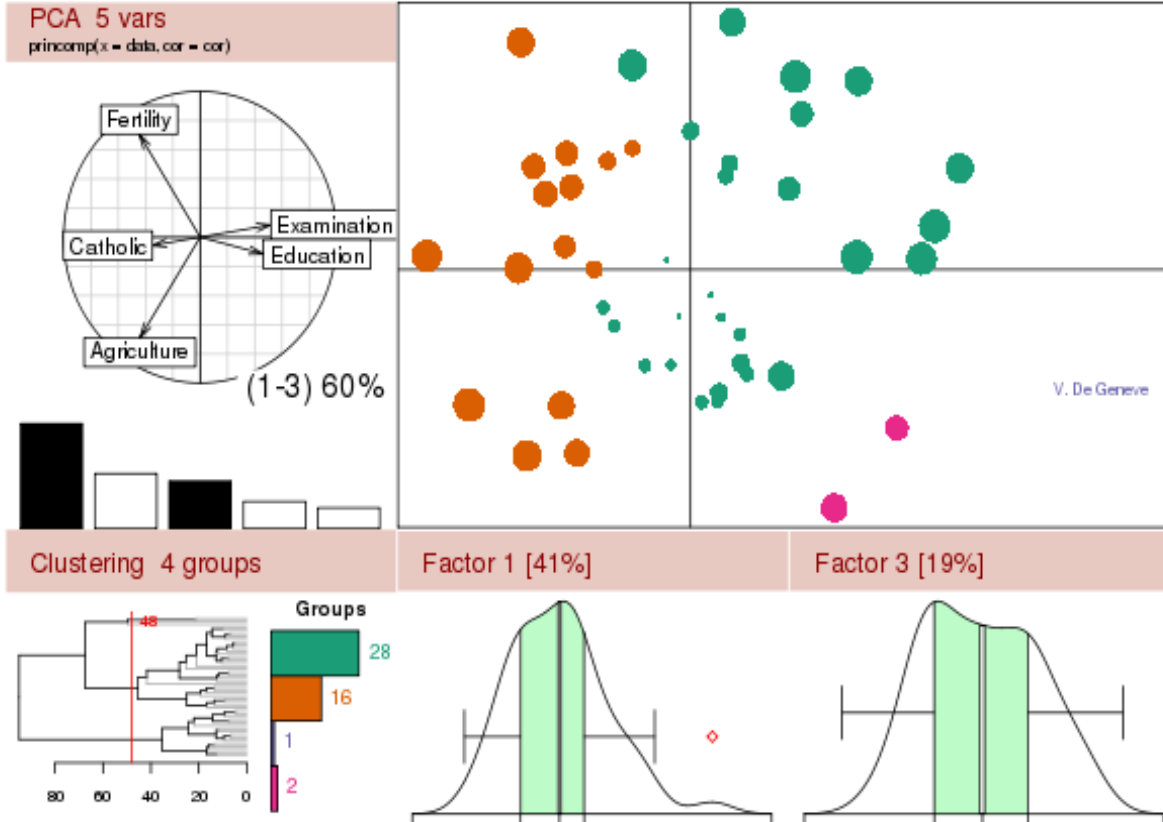
Opened nodes (19); Saved nodes(1)

Big Data Analytics

Real Time Analytics Platform



R Statistical Programming Language



Open source language and environment

Used for statistical computing and graphics

Strength in easily producing publication-quality plots

Highly extensible with open source community R packages

Drive Value from Big Data

Conclusions



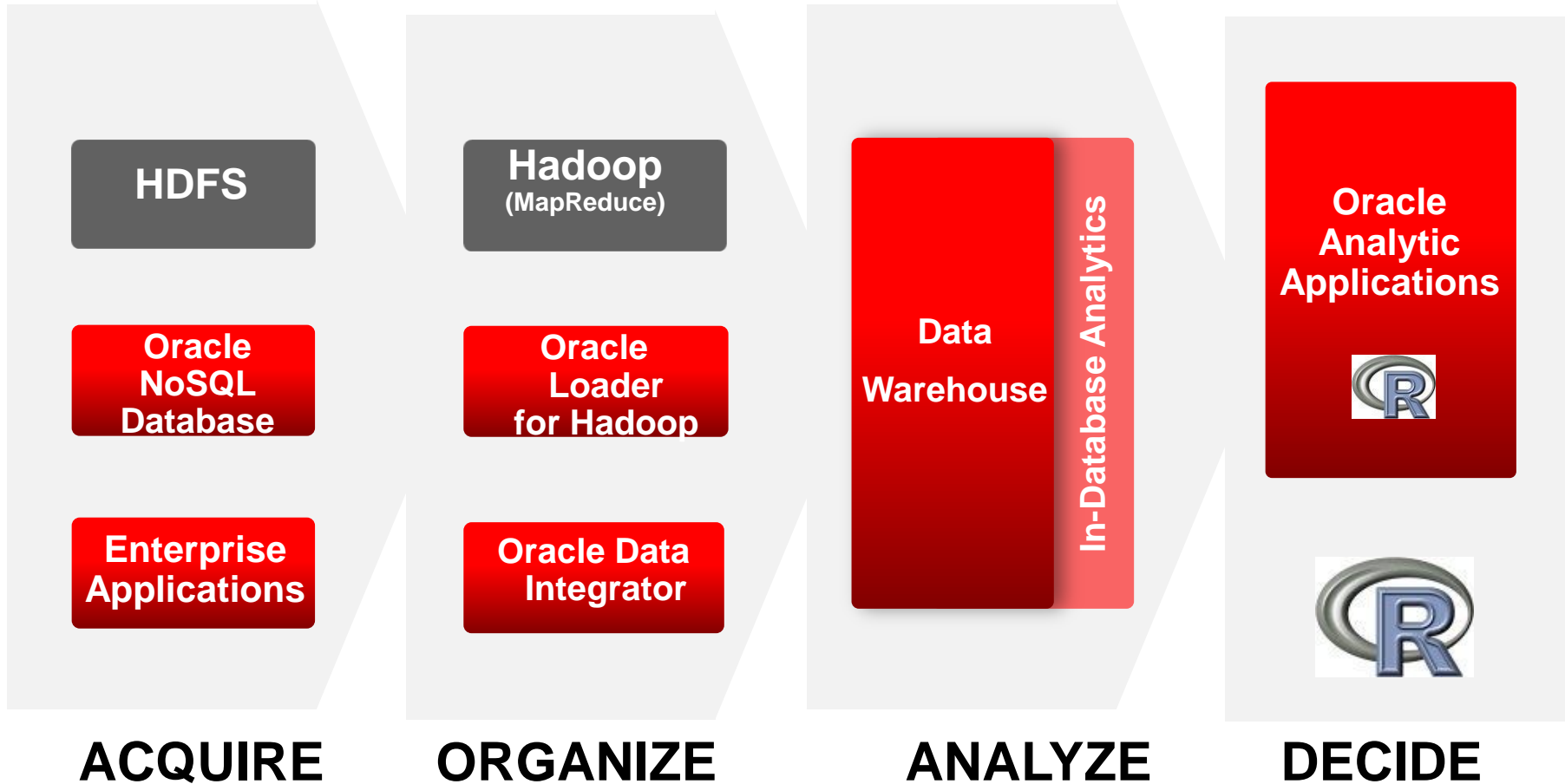
Big Data Appliance

Big Data for the Enterprise

- Optimized and Complete
 - Everything you need to store and integrate your lower information density data
- Integrated with Oracle Exadata
 - Analyze all your data
- Easy to Deploy
 - Risk Free, Quick Installation and Setup
- Single Vendor Support
 - Full Oracle support for the entire system and software set



Oracle Integrated Solution Stack for Big Data



Oracle: Big Data for the Enterprise

- The most comprehensive solution
 - Includes everything needed to acquire, organize and analyze all your data
- Optimized for Extreme Analytics
 - Deepest analytics portfolio with access to all data
- Engineered to Work Together
 - Eliminate deployment risk and support risk
- Enterprise Ready
 - Deliver extreme performance and scalability

Questions



ORACLE®