




ORACLE®

Introduction to Big Data Technology



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.



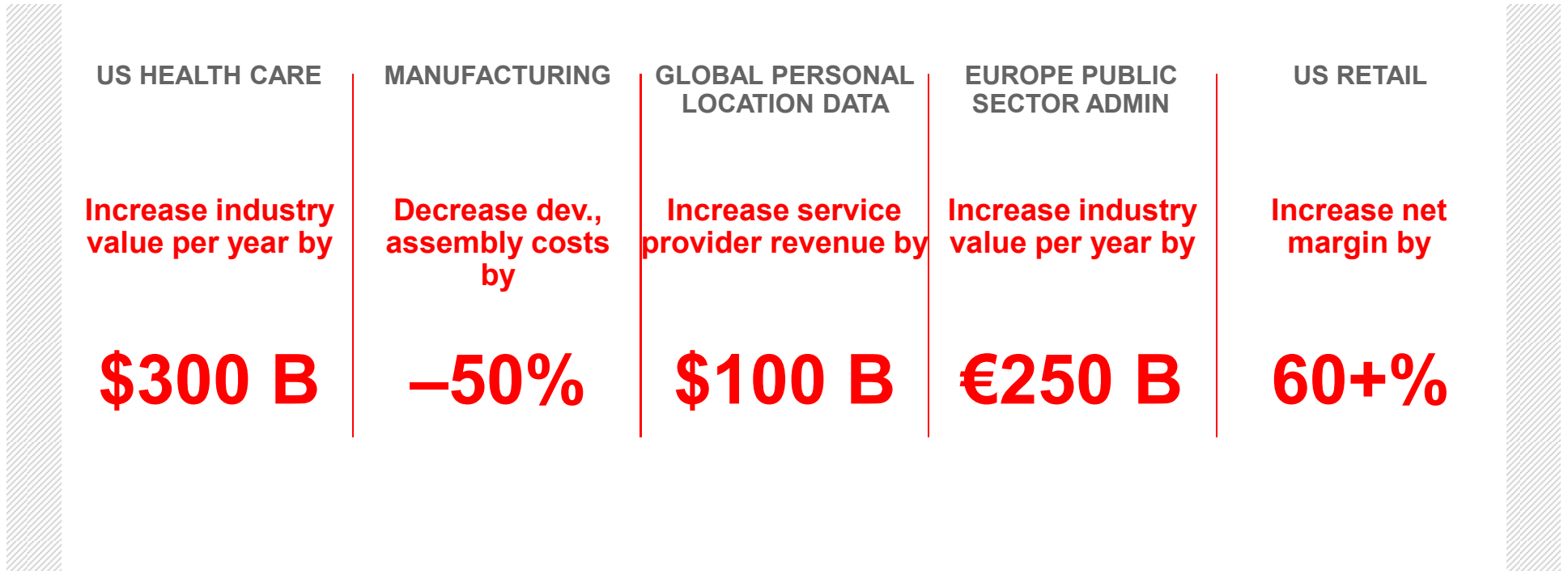
Agenda

- What is Big Data
- Real World Use of Big Data
- Architecture Overview
- Oracle Big Data Solution
- Summary



WHAT IS BIG DATA?





Why Is Big Data Important?



ORACLE

Source: * McKinsey Global Institute: Big Data – The next frontier for innovation, competition and productivity (May 2011)

Data are Growing

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

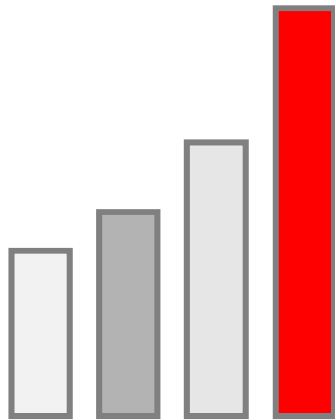


What is Big data

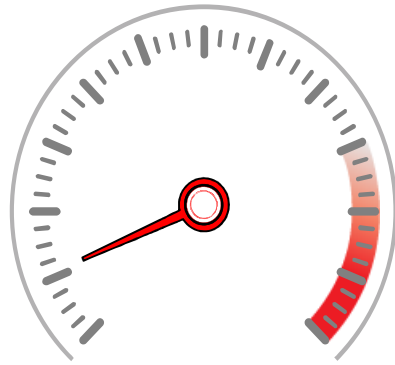
- Big data analytics is concerned with the analysis of **large volumes** of transaction/event data and **behavioral analysis** of human/human and human/system interactions. (Gartner)
- Big data represents the **collection of technologies** that handle **large data volumes** well beyond that inflection point and for which, at least in theory, hardware resources required to manage data by volume track to an almost straight line rather than a curve. (IDC)



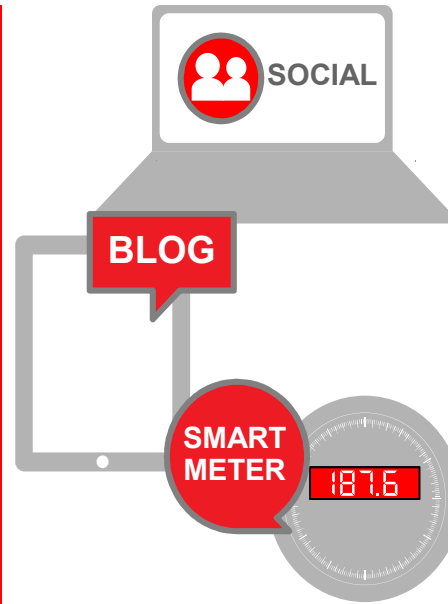
What is Big data



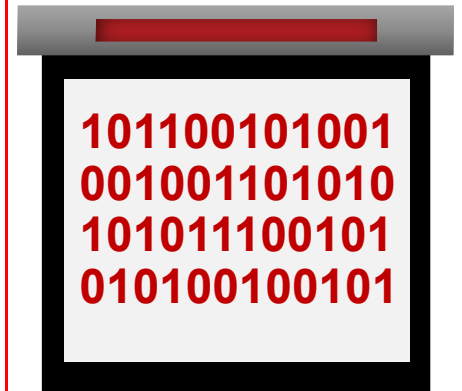
VOLUME



VELOCITY



VARIETY

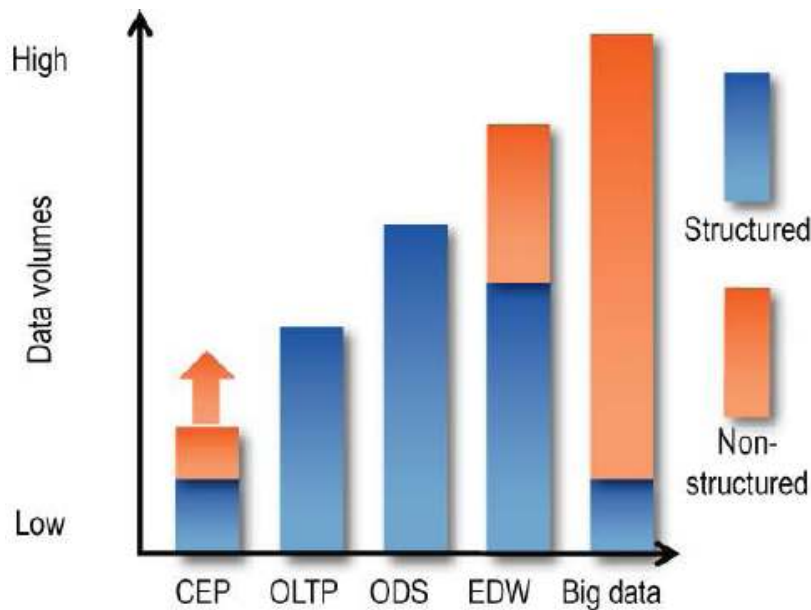


VALUE

Structured Non Structured

Level	Example
Structured	Relational database
Semi-structured	XML data files
Quasi-structured	Text documents
Unstructured	Images and video

Non Structured



© 2011 Gartner, Inc. and/or its affiliates. All rights reserved.

A new class of problems has emerged which demands an ability to accept and manage data **without advanced knowledge of its structure or format.**

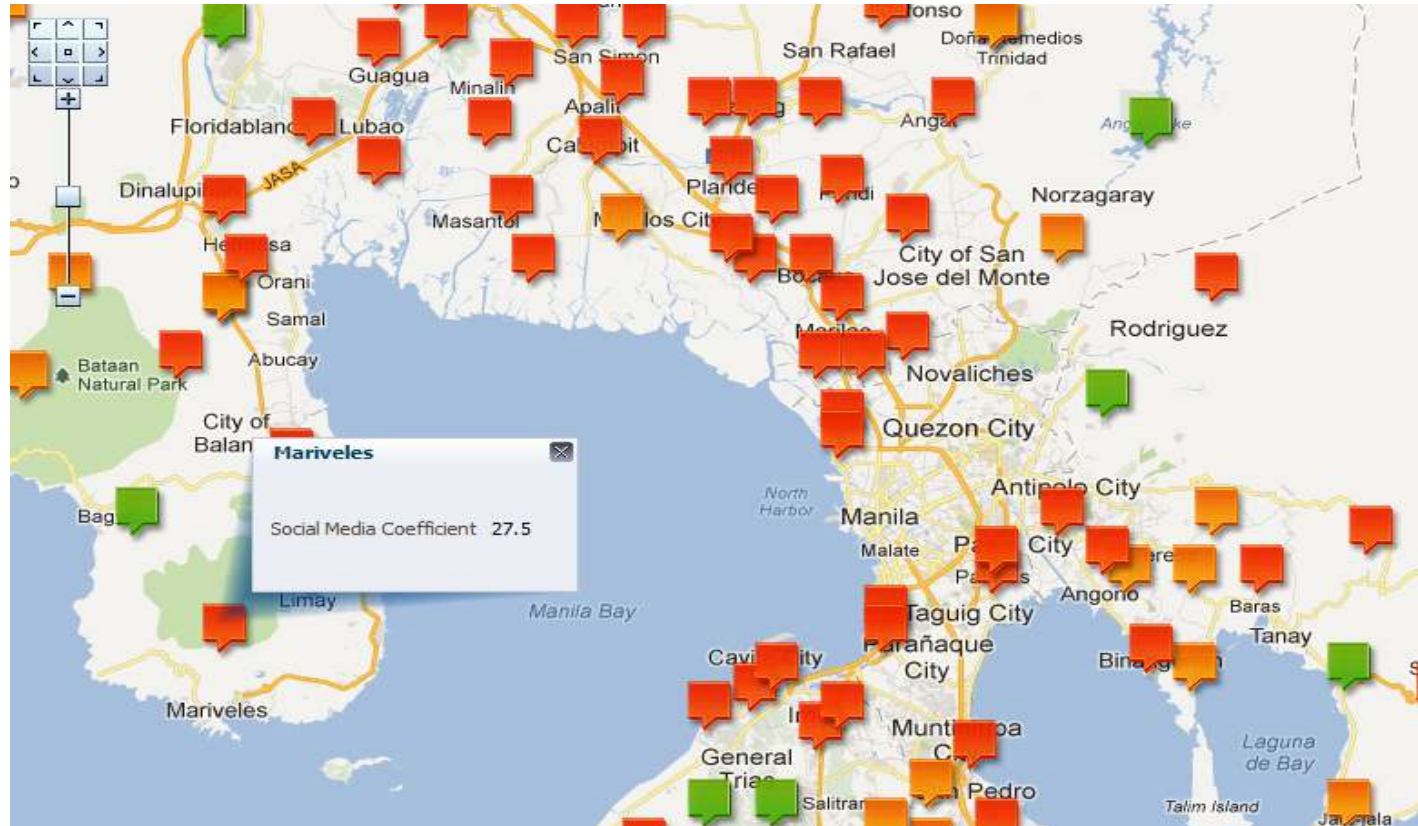


Real world use of big data

Today's Challenge	New Data	What's Possible
Healthcare Expensive office visits	Remote patient monitoring	Preventive care, reduced hospitalization
Manufacturing In-person support	Product sensors	Automated diagnosis, support
Location-Based Services Based on home zip code	Real time location data	Geo-advertising, traffic, local search
Public Sector Standardized services	Citizen surveys	Tailored services, cost reductions
Retail One size fits all marketing	Social media	Sentiment analysis segmentation

Public Healthcare

Management of Outbreak Through Early Detection of Clusters





Government Officer

2. Government creates/ revises National Health Program



Doctor



Nurse



Officer

3. Public/Private Hospital executes Health Program integrated EHR/EMR Systems



Cloud

Big Data

Suggested Health Improvement (Secured Personal Access)

Personal Health Improvement

1. Blood Pressure Coach Tracking

4. Health Tracking Health check up records

1. Blood Pressure Sleep Tracking

4. Health Tracking Health check up records

1. Blood Pressure Heart rate Tracking

4. Health Tracking Health check up records



Integrated Medical Device



Integrated Medical Device



Integrated Medical Device



Integrated Medical Device

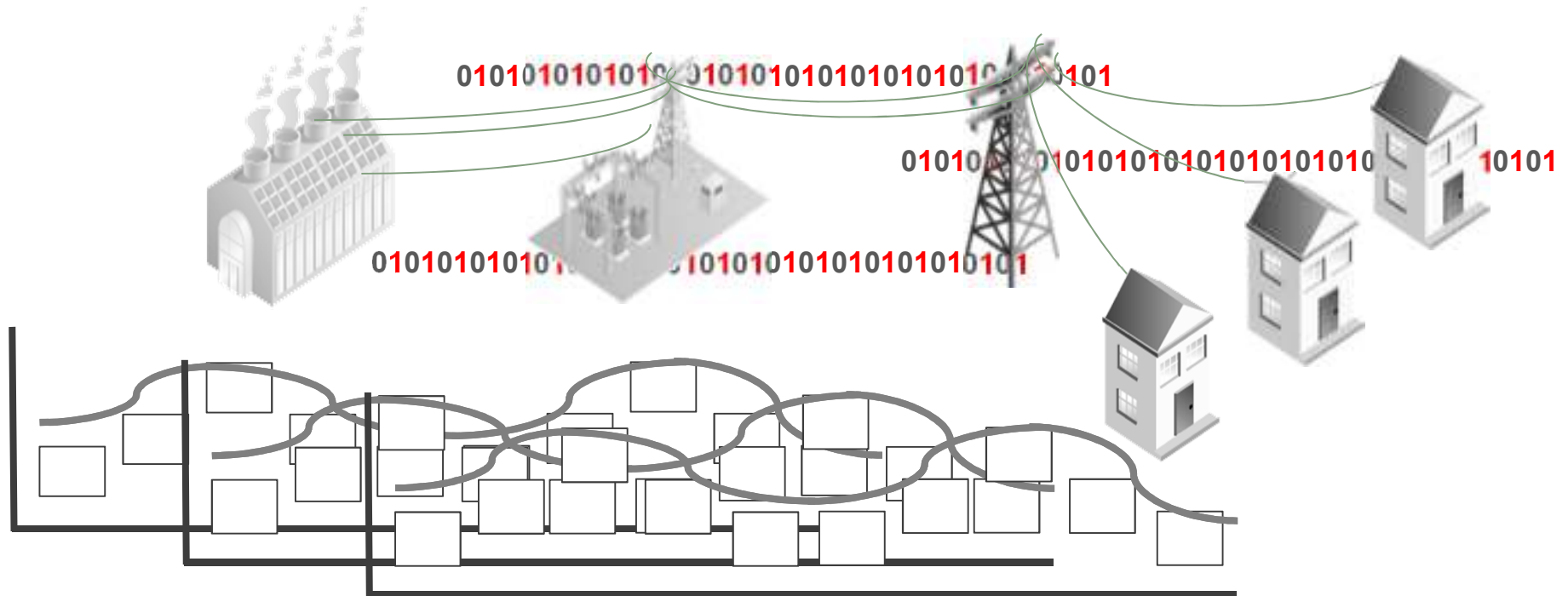


Integrated Medical Device

ORACLE

What does a Big Data World look like?

Utilities



What they collect

Smart Metering - *Monitors power usage*

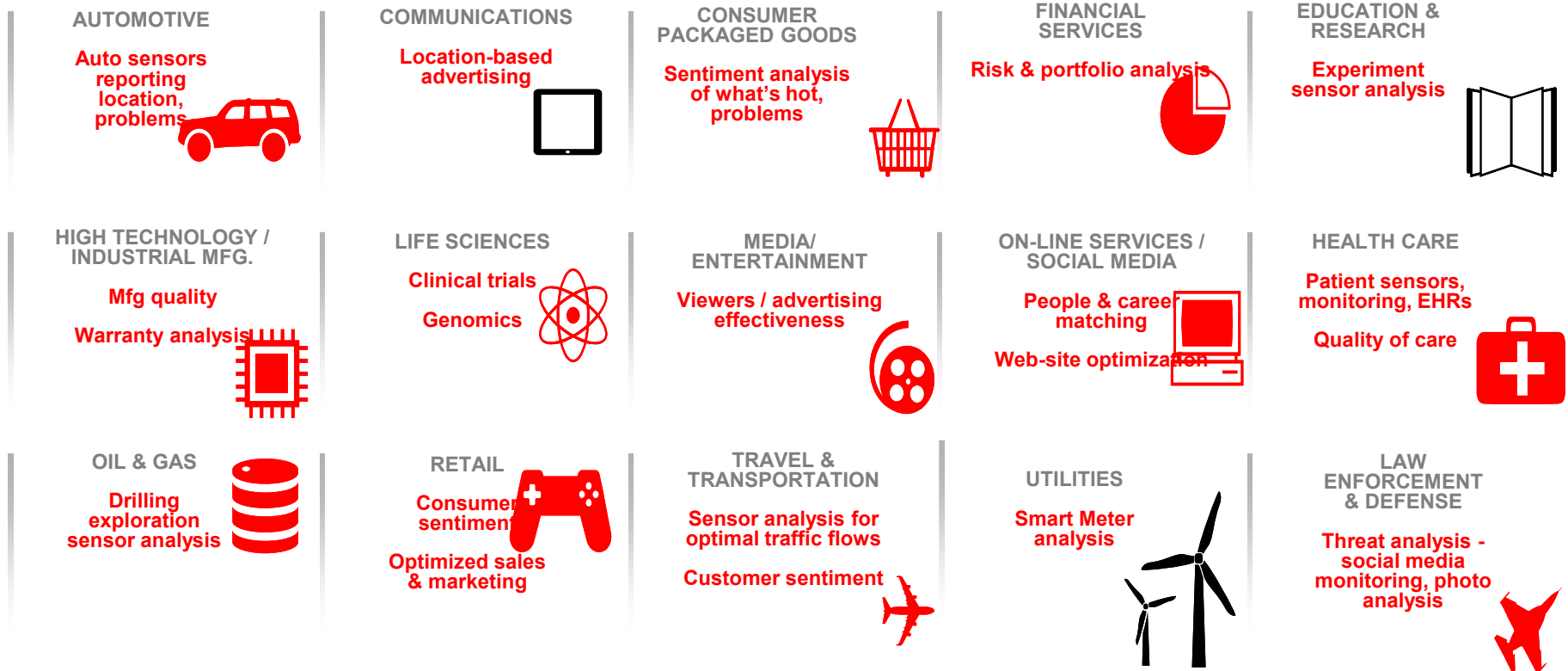
How they use it

Better demand planning
Better targeted marketing
Better targeted products based on individuals power needs

Big Data means...

The ability to predict demand at household level
Reduce exposure to spot market

Real world use of big data



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



ARCHITECTURE OVERVIEW

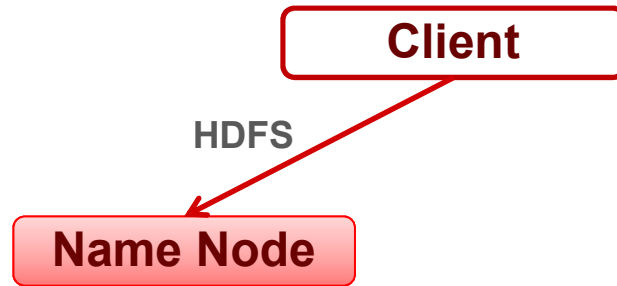


Basic Architecture

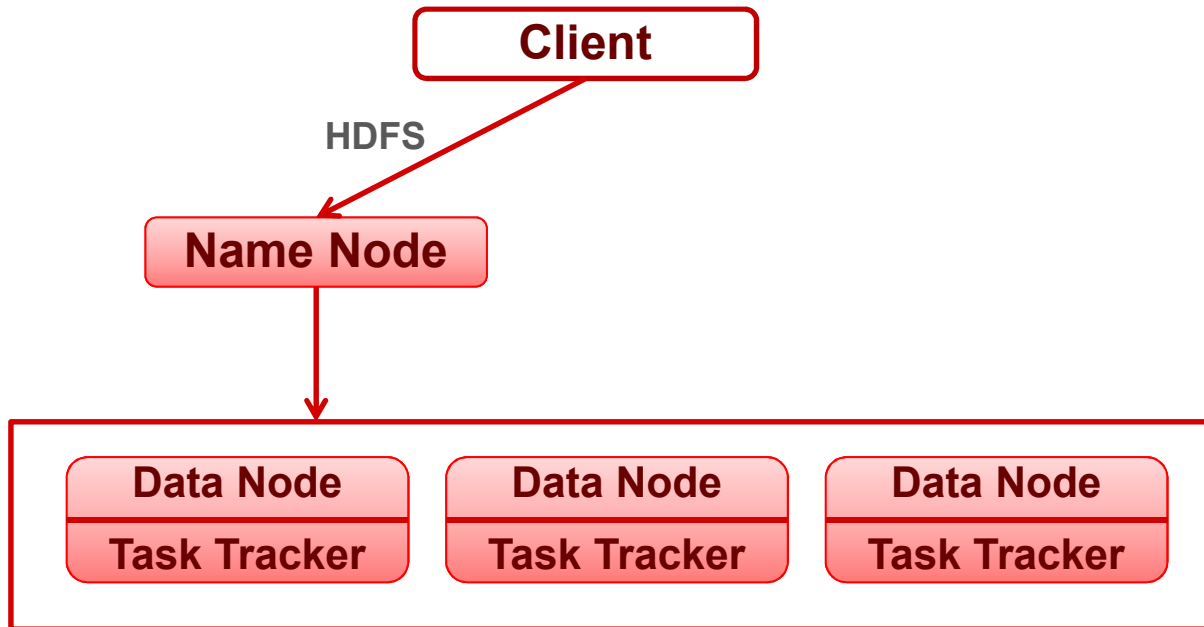
Client



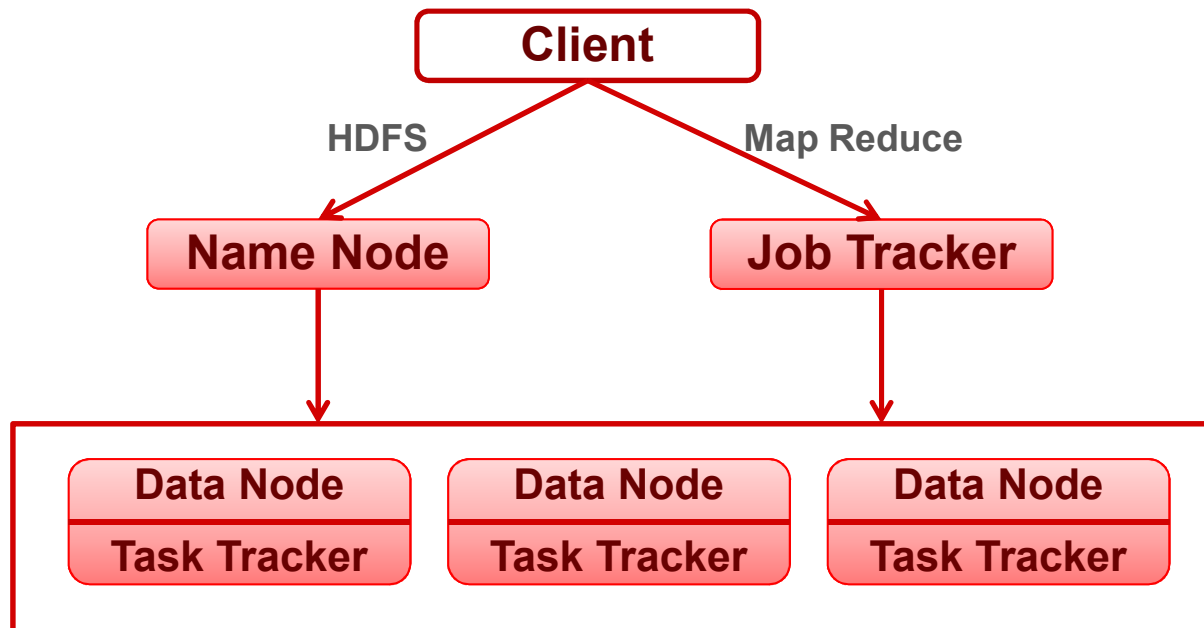
Basic Architecture



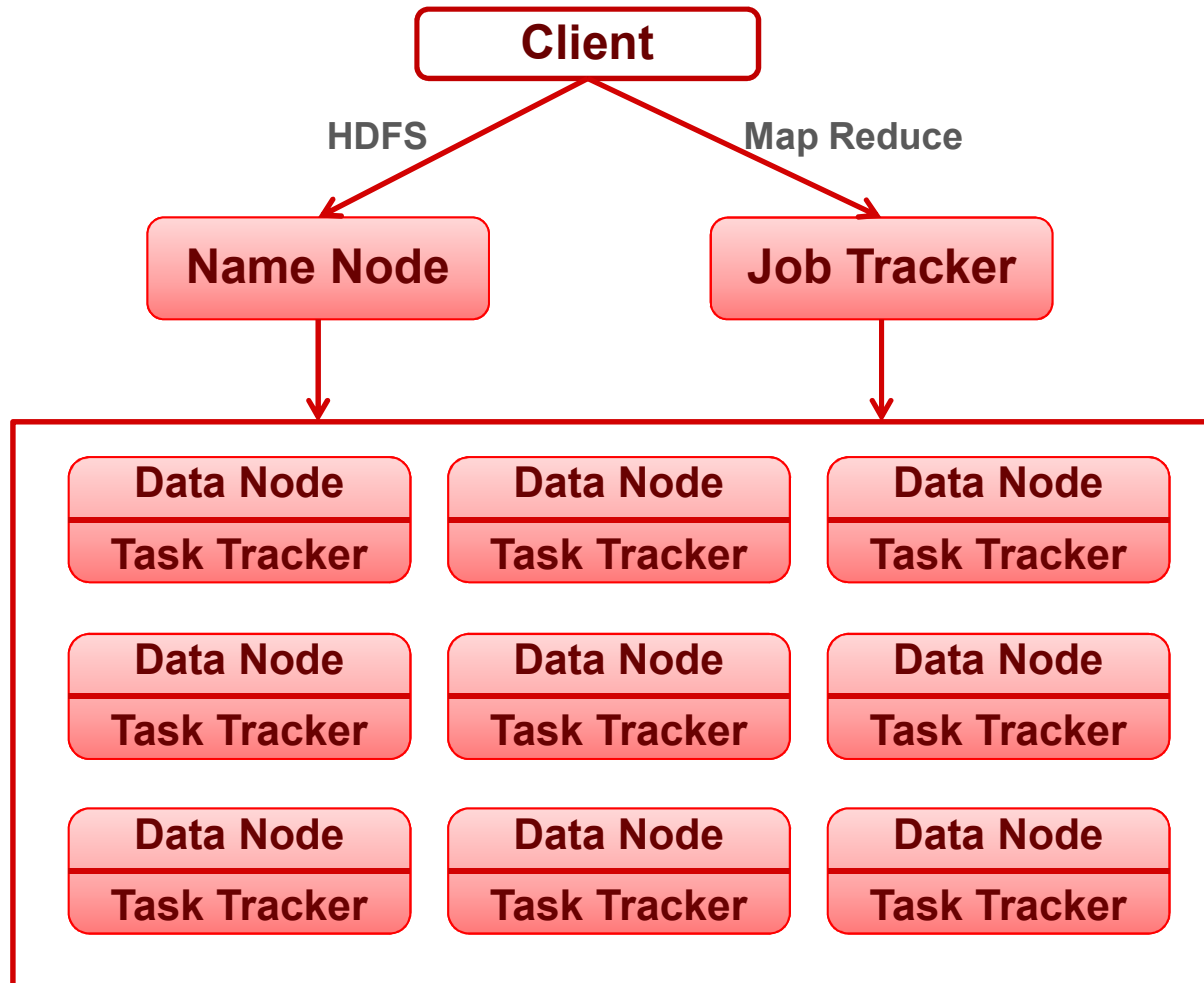
Basic Architecture



Basic Architecture



Basic Architecture



The concept is not new ...



The concept in not new ...



The concept is not new ...



The concept is not new ...





The concept is not new ...



The concept is not new ...



The concept is not new ...



The concept is not new ...





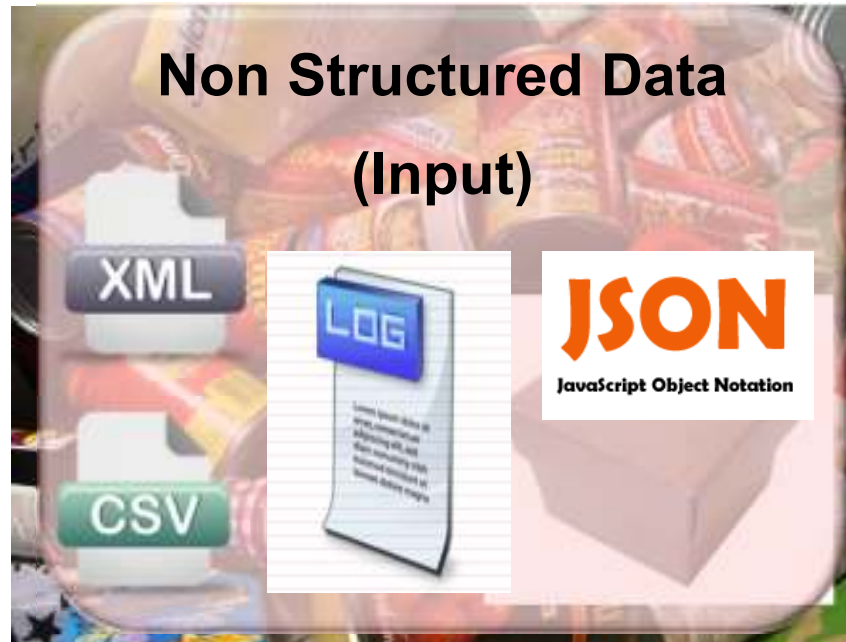
Mapping to world of Big data

Data



Mapping to world of Big data

Data



Mapping to world of Big data

Data Integration



Mapping to world of Big data

Data Integration





Mapping to world of Big data

Data Management



Mapping to world of Big data

Data Management



Mapping to world of Big data

Data Processing



Mapping to world of Big data

Data Processing



Mapping to world of Big data

Data Query



Mapping to world of Big data

Data Query



Mapping to world of Big data Analysis



Mapping to world of Big data Analysis

Analysis (statistical computing and graphics env)



Analysis (machine learning and data mining library)



Analysis (displaying, monitoring & analyzing)



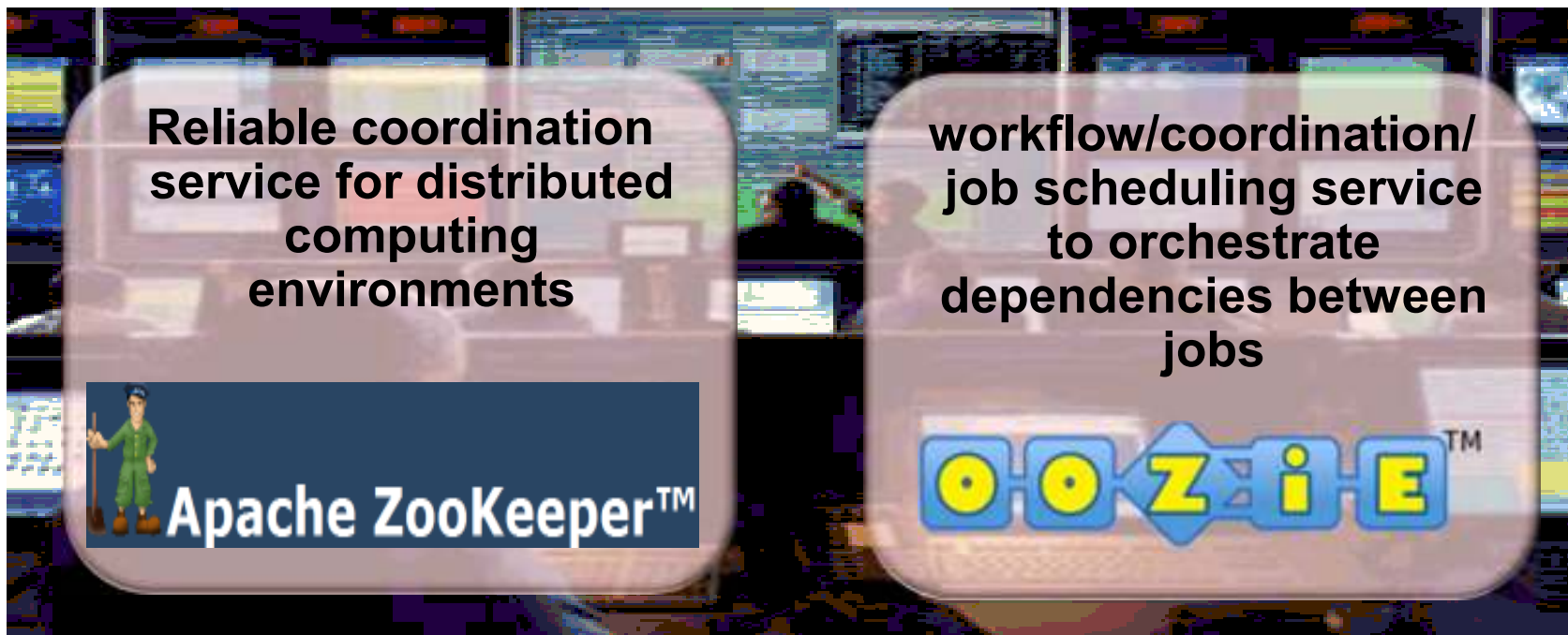
Mapping to world of Big data

Management and Co-ordination



Mapping to world of Big data

Management and Co-ordination



**Reliable coordination
service for distributed
computing
environments**

Apache ZooKeeper™

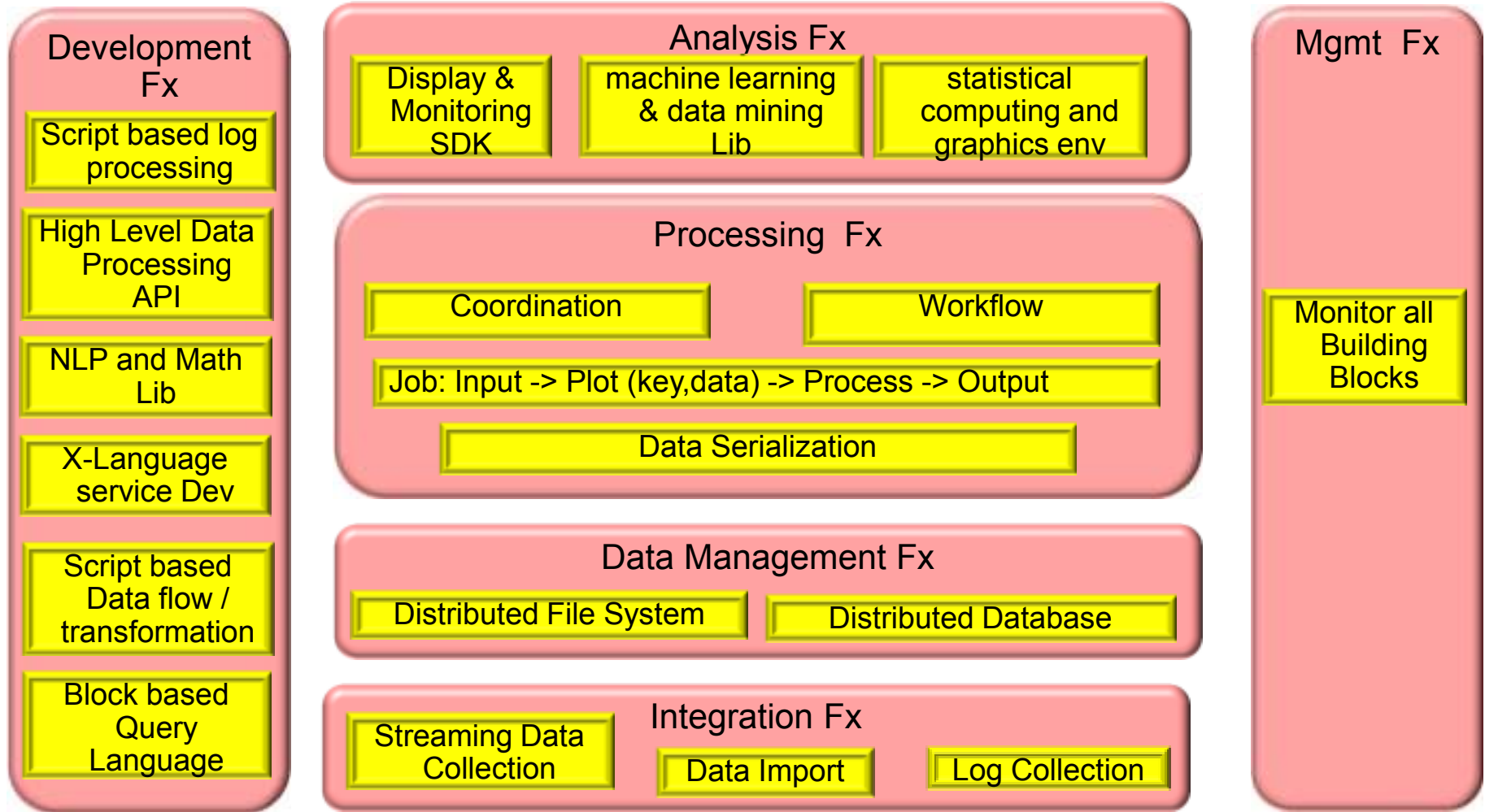
**workflow/coordination/
job scheduling service
to orchestrate
dependencies between
jobs**

OOZIE™



Big Data Functional Building Blocks

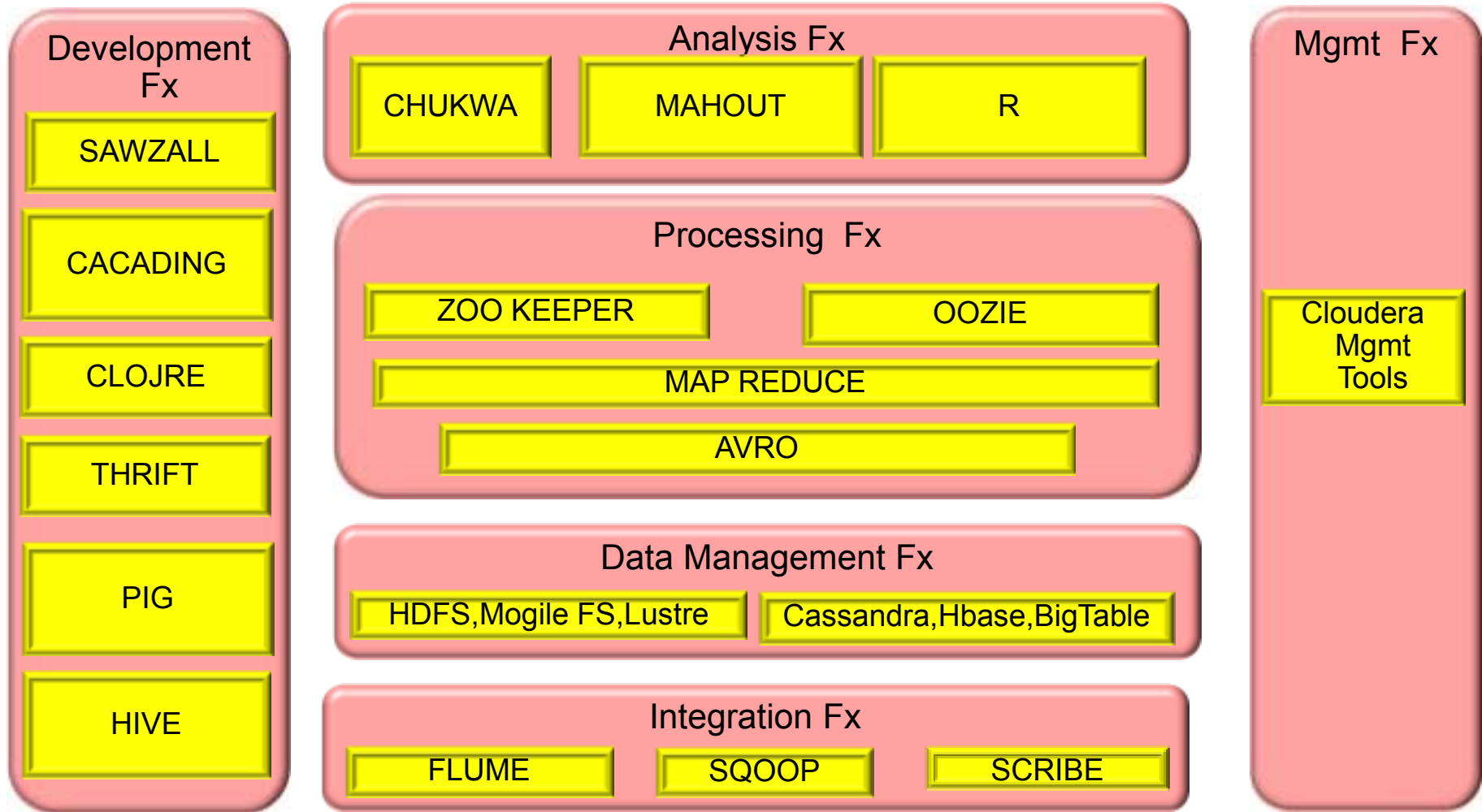
Ecosystem





Big Data Solution Building Blocks

Technologies





ORACLE BIG DATA SOLUTION

Oracle Big Data Appliance

Big Data for the Enterprise

HDFS ✓

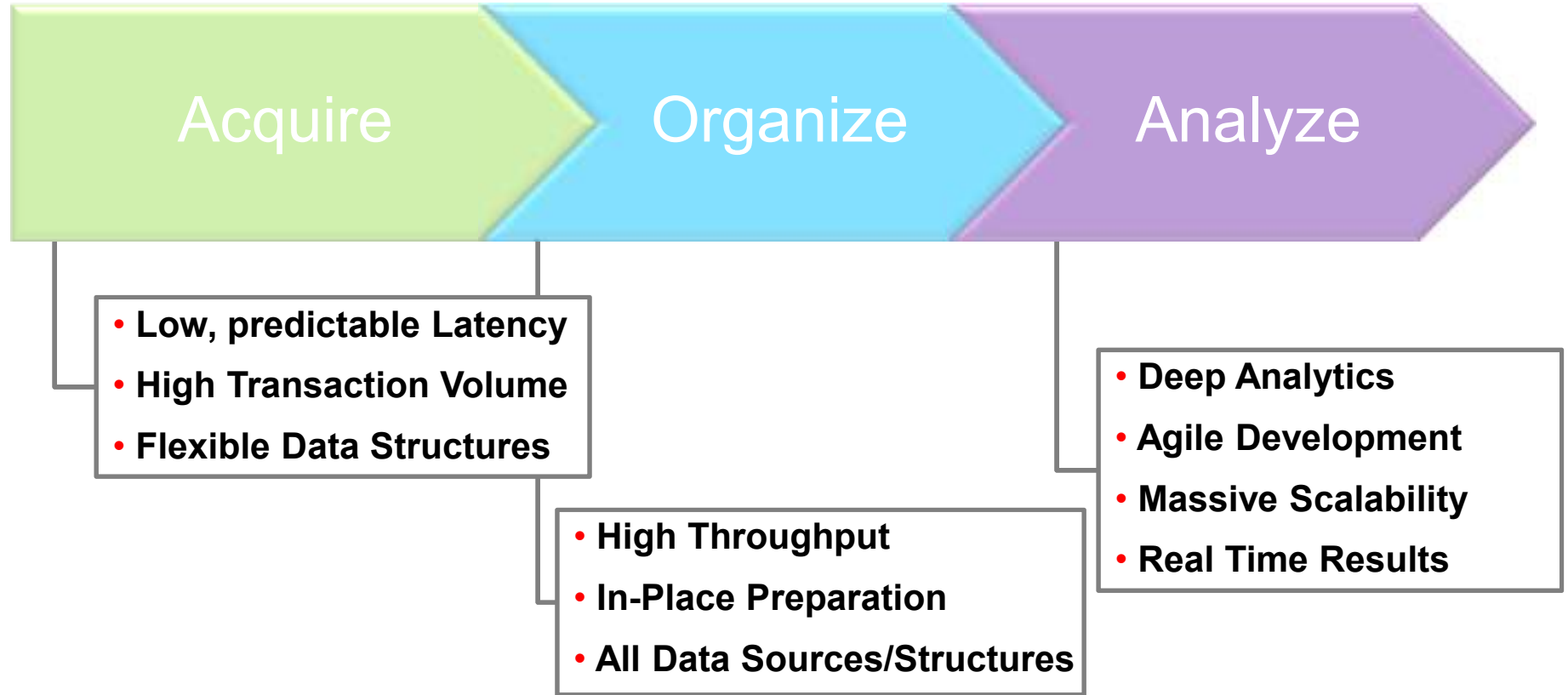
Map Reduce ✓

Ecosystems ✓

Enterprise Support ✓

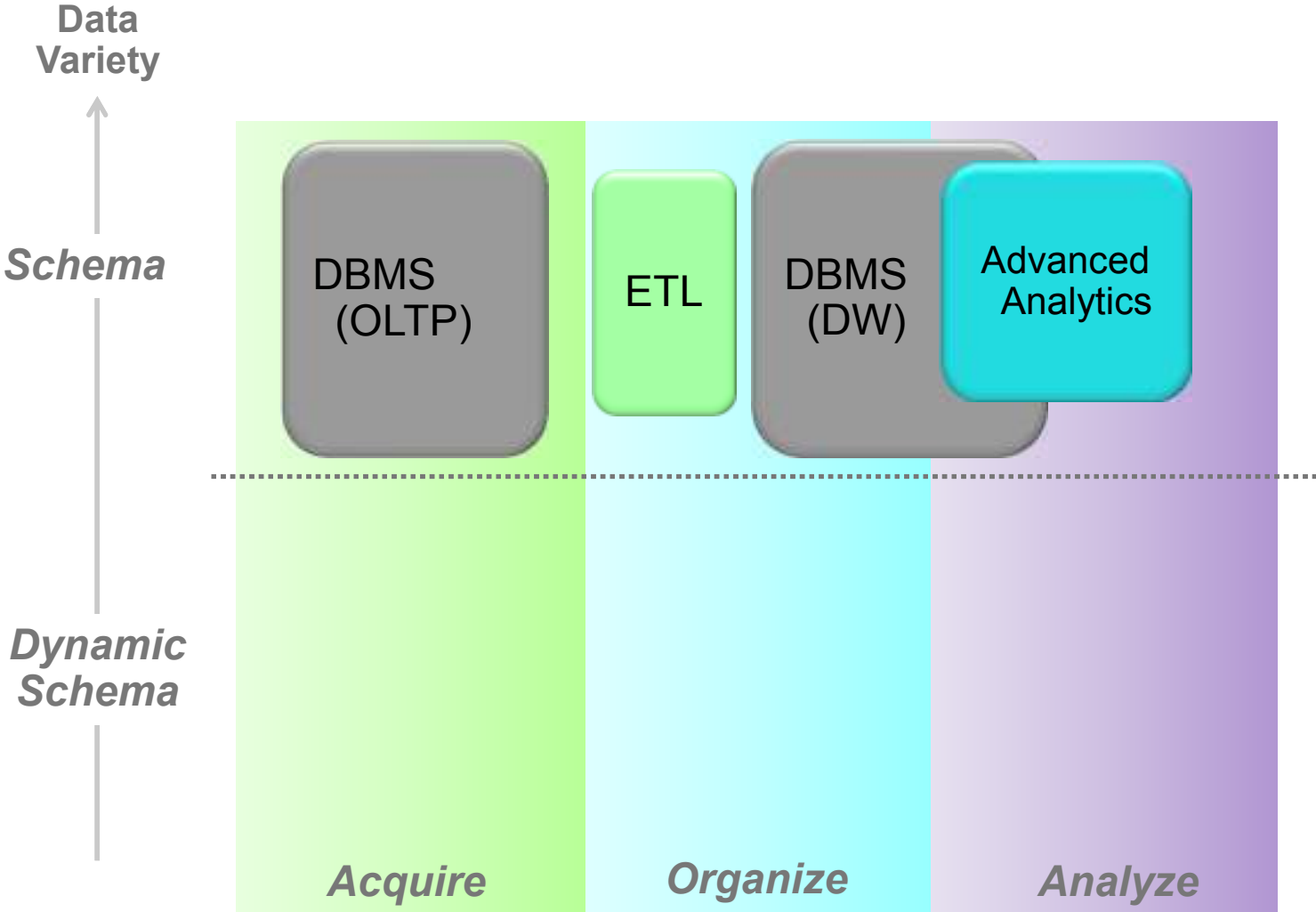


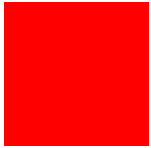
Big Data: Infrastructure Requirements



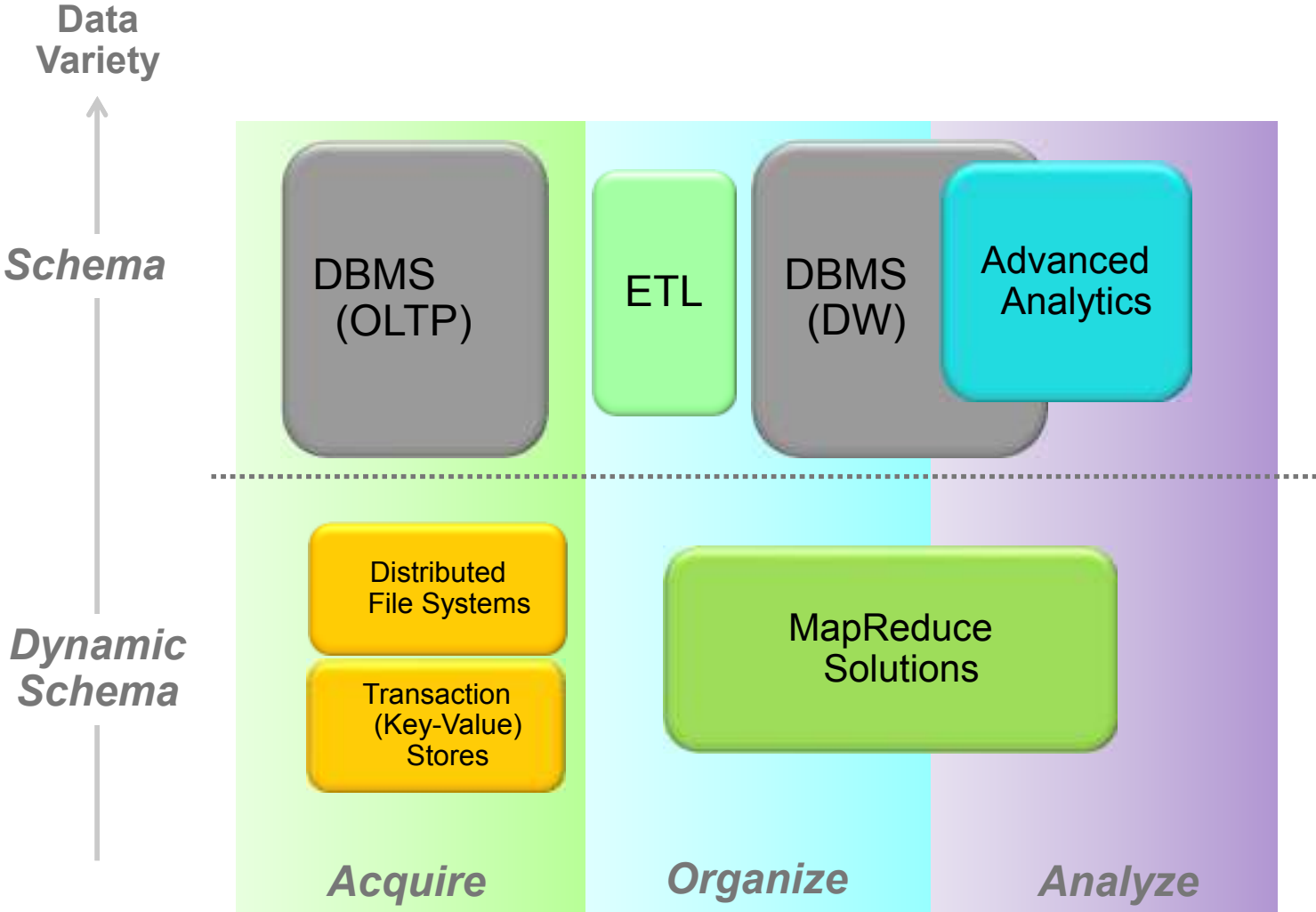


Divided Solution Spectrum



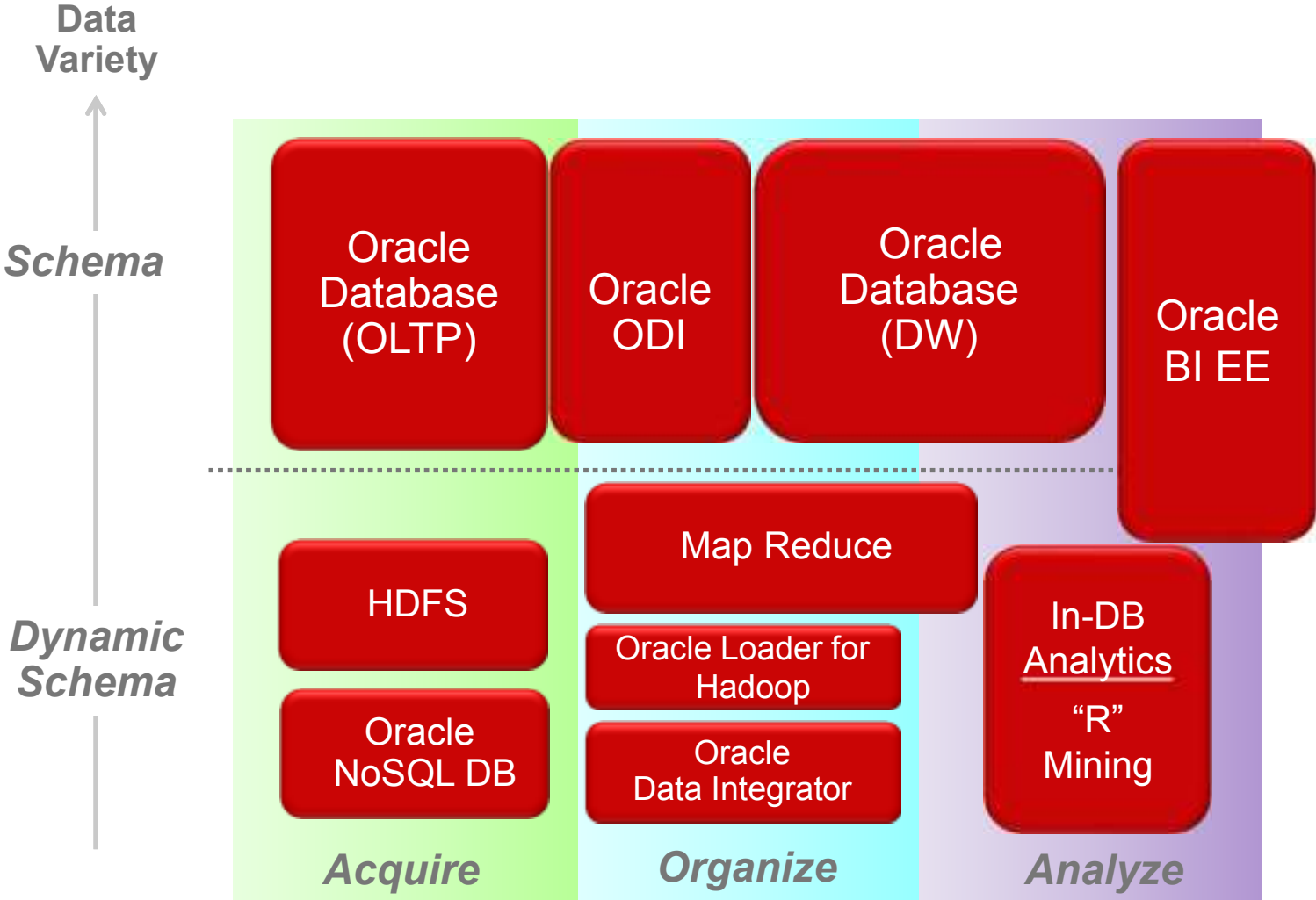


Divided Solution Spectrum



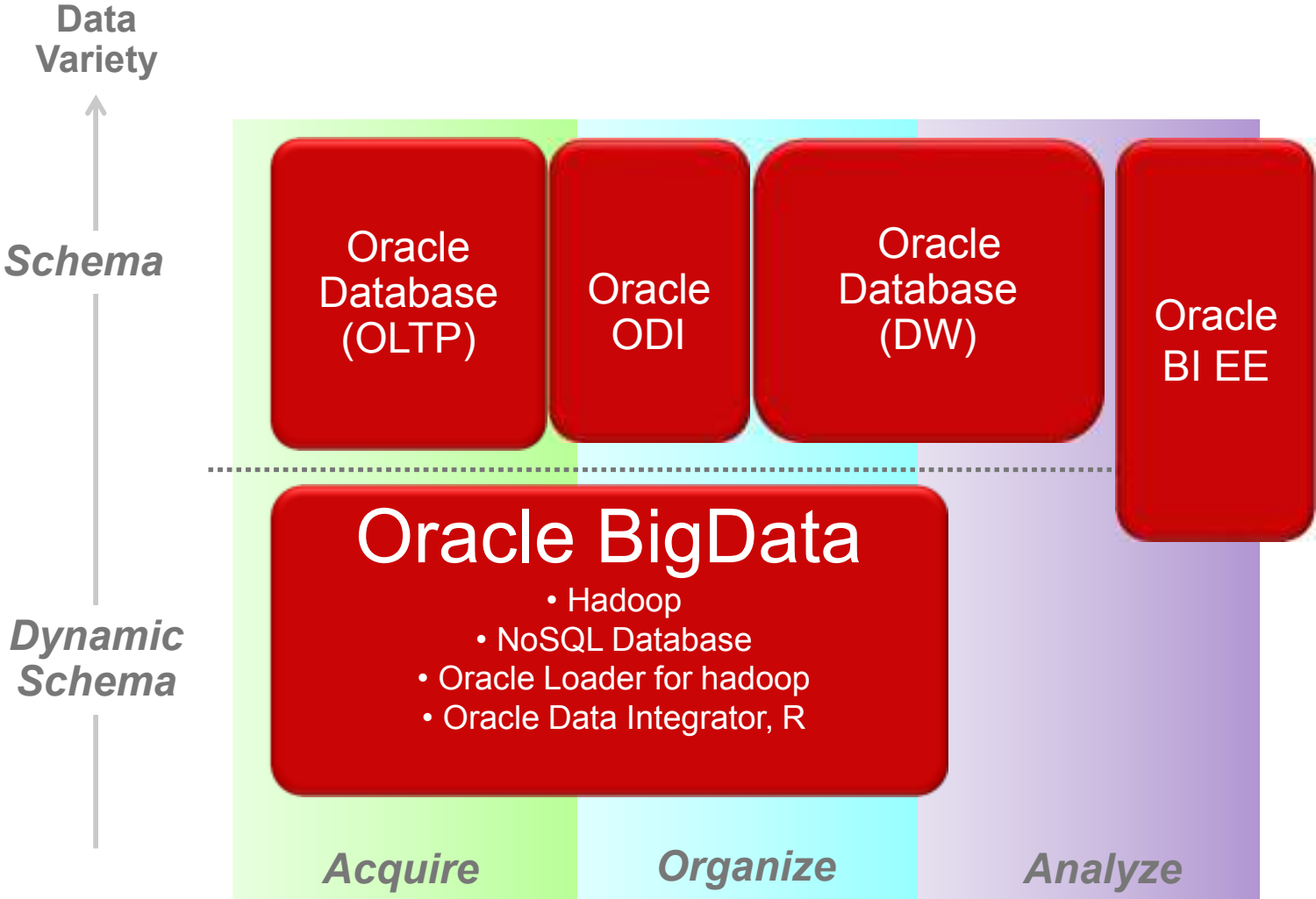


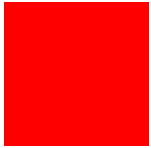
Divided Solution Spectrum



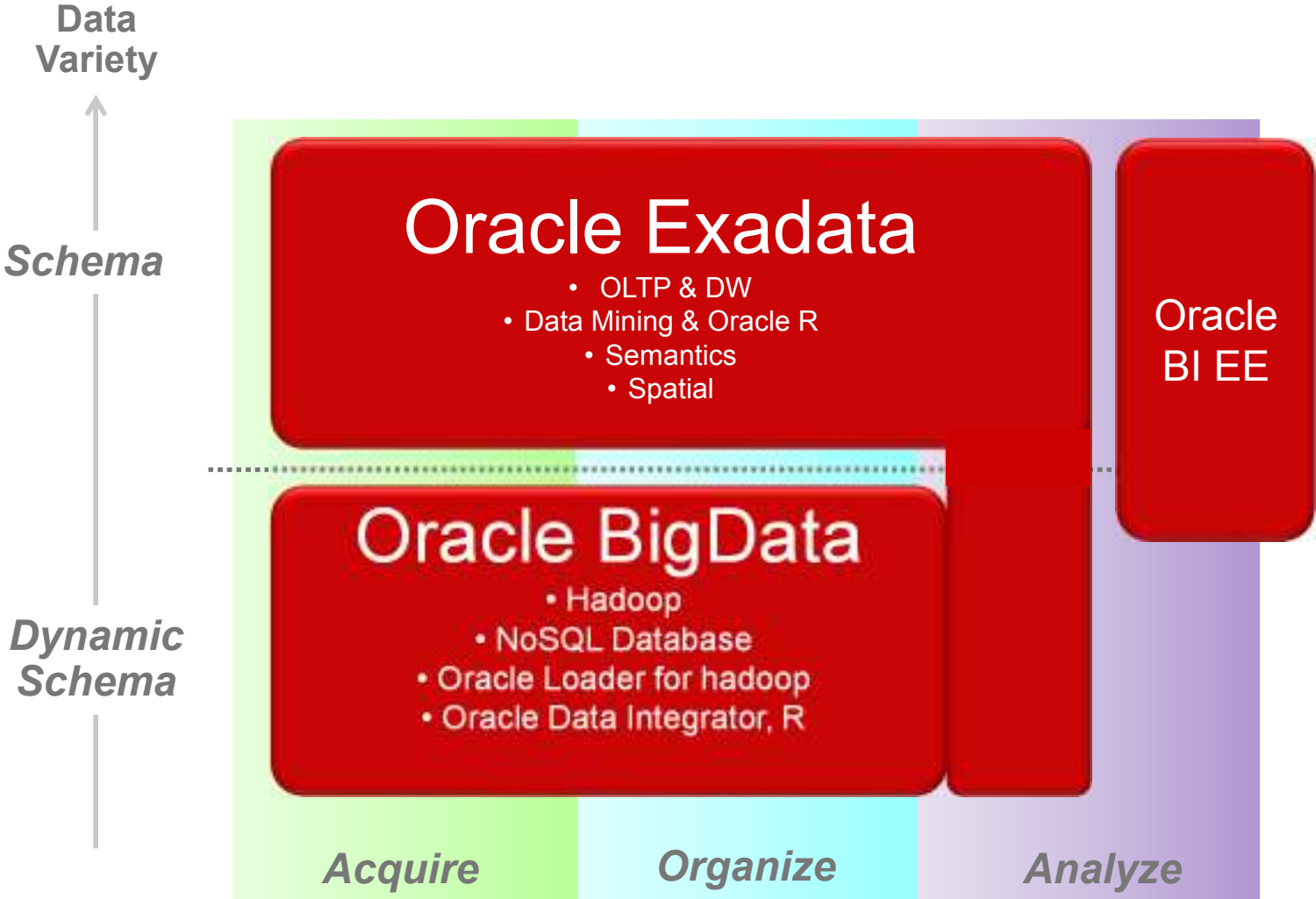


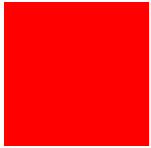
Divided Solution Spectrum



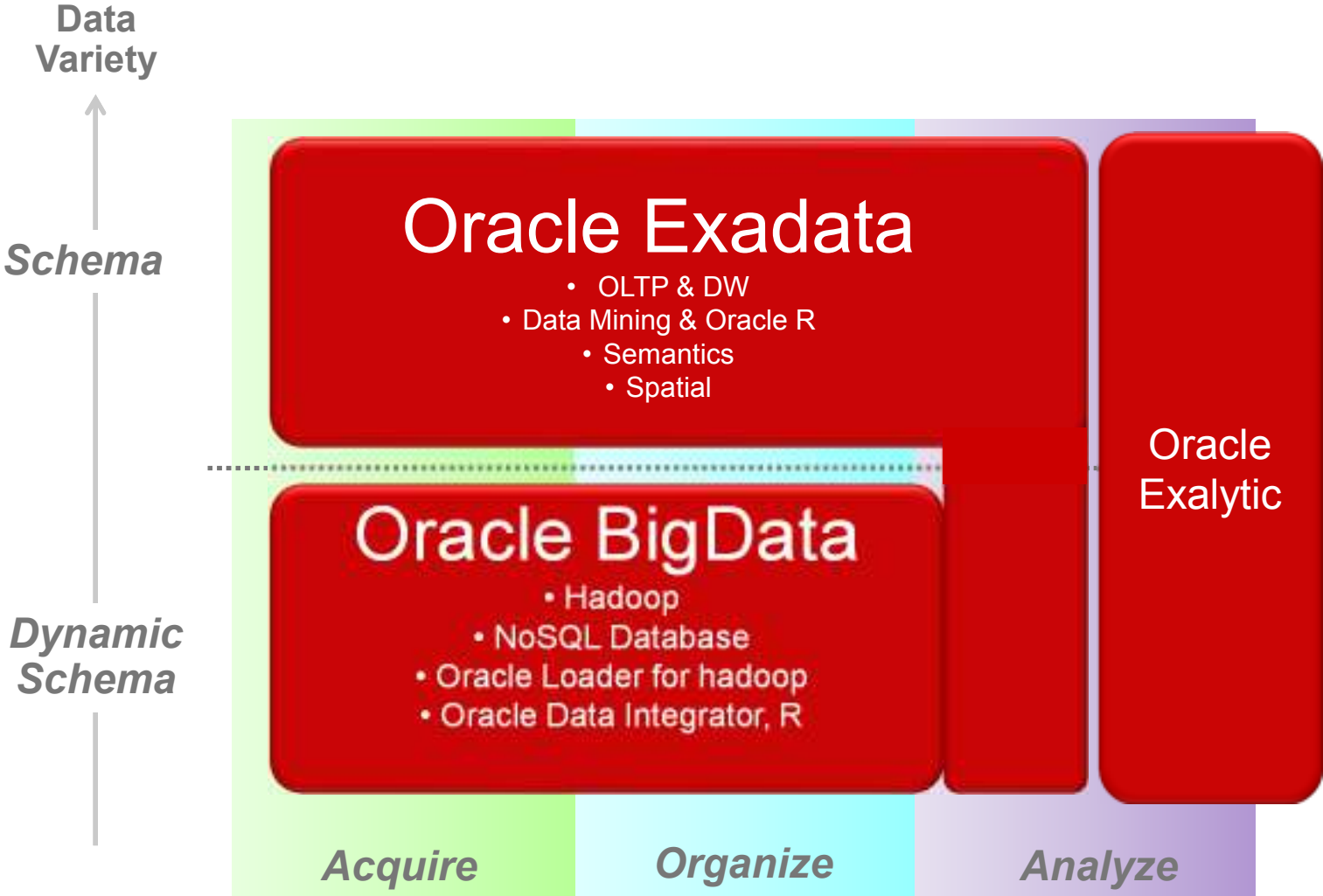


Divided Solution Spectrum



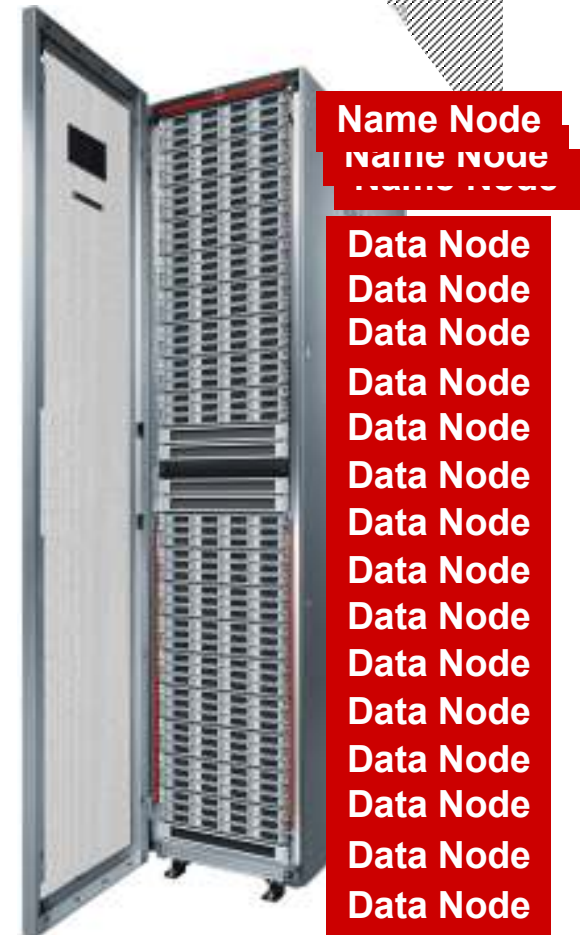


Divided Solution Spectrum



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

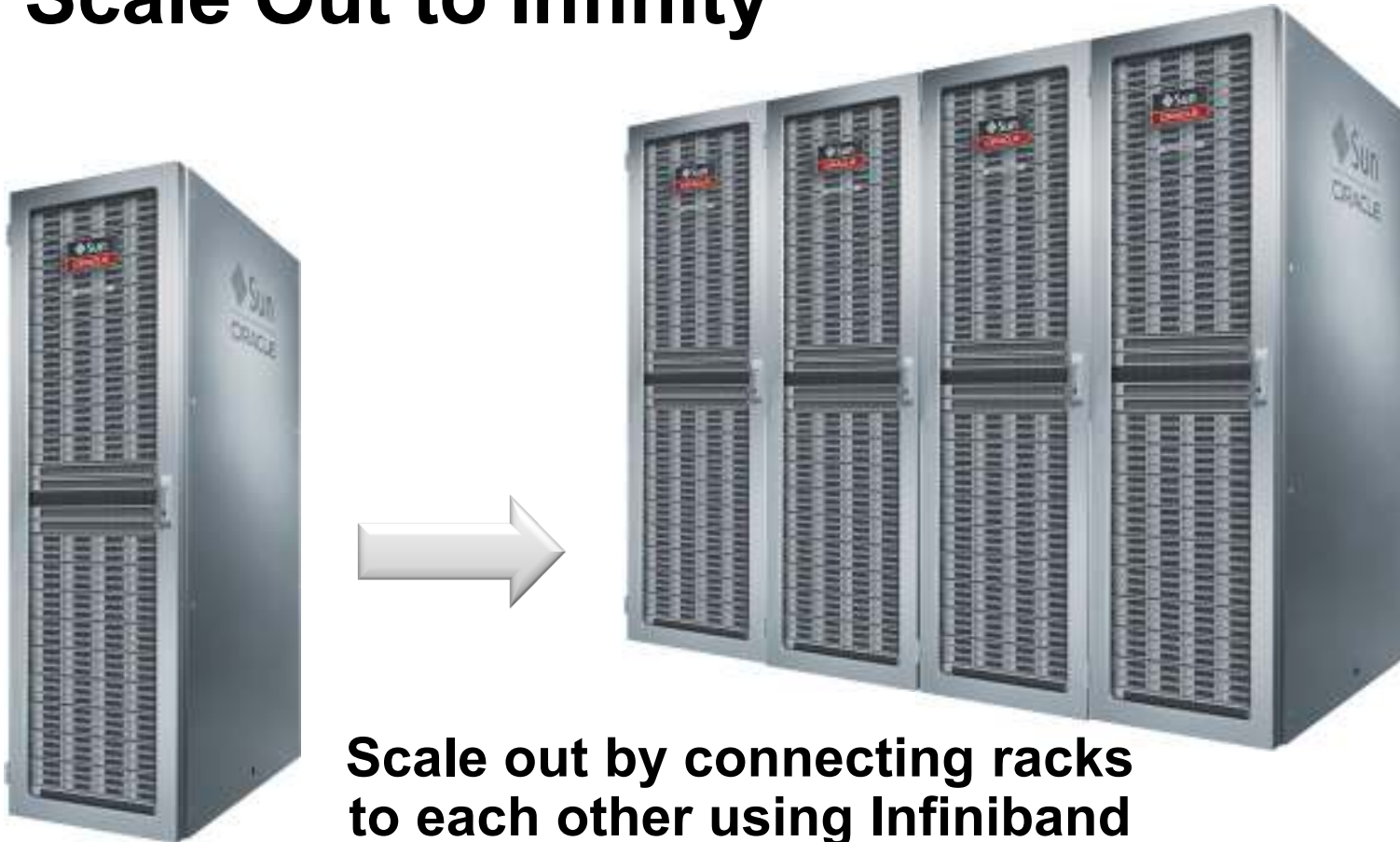


Oracle Big Data Appliance Software

- Oracle Linux 5.6
- Apache Hadoop Distribution
- R Distribution
- Oracle NoSQL Enterprise Edition
- Oracle Data Integrator Application
Adapter for Hadoop
- Oracle Loader for Hadoop

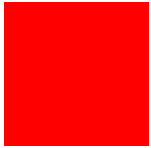


Scale Out to Infinity



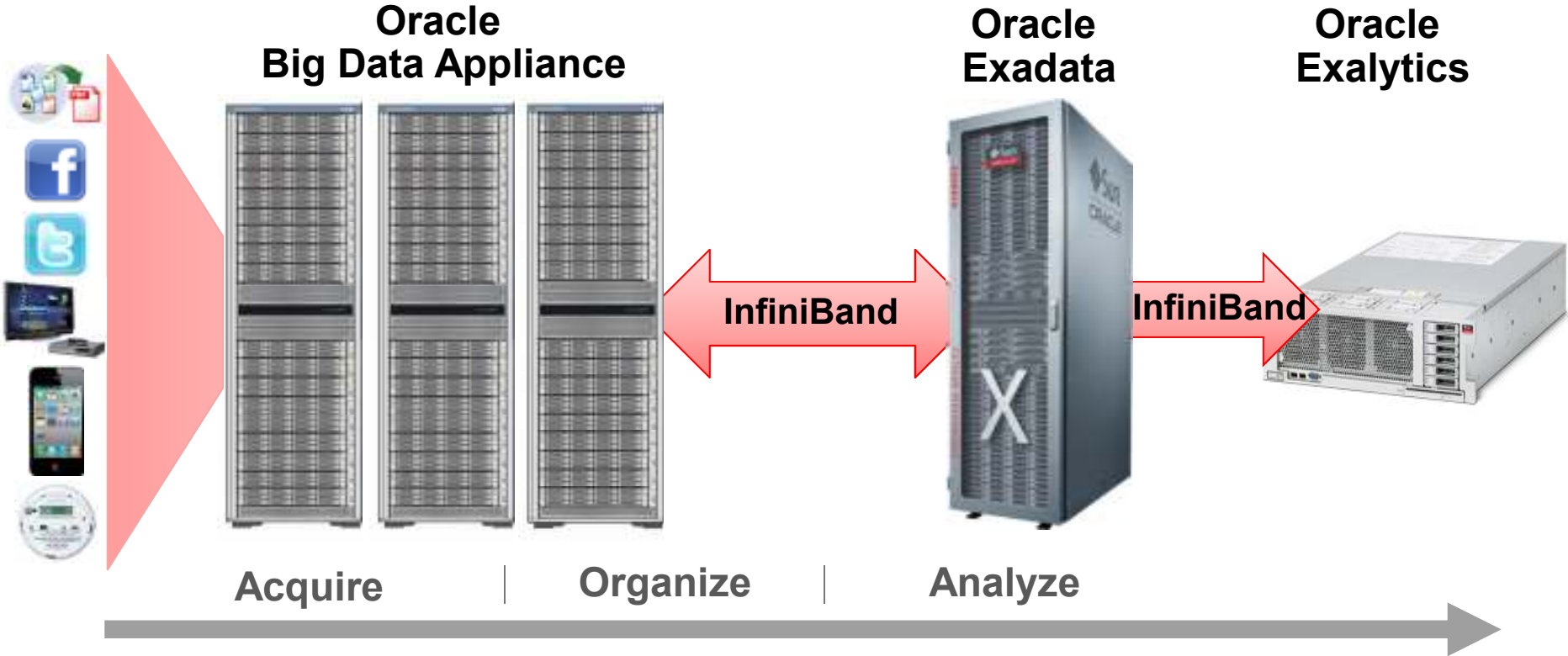
Scale out by connecting racks to each other using Infiniband

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



Big Data Appliance

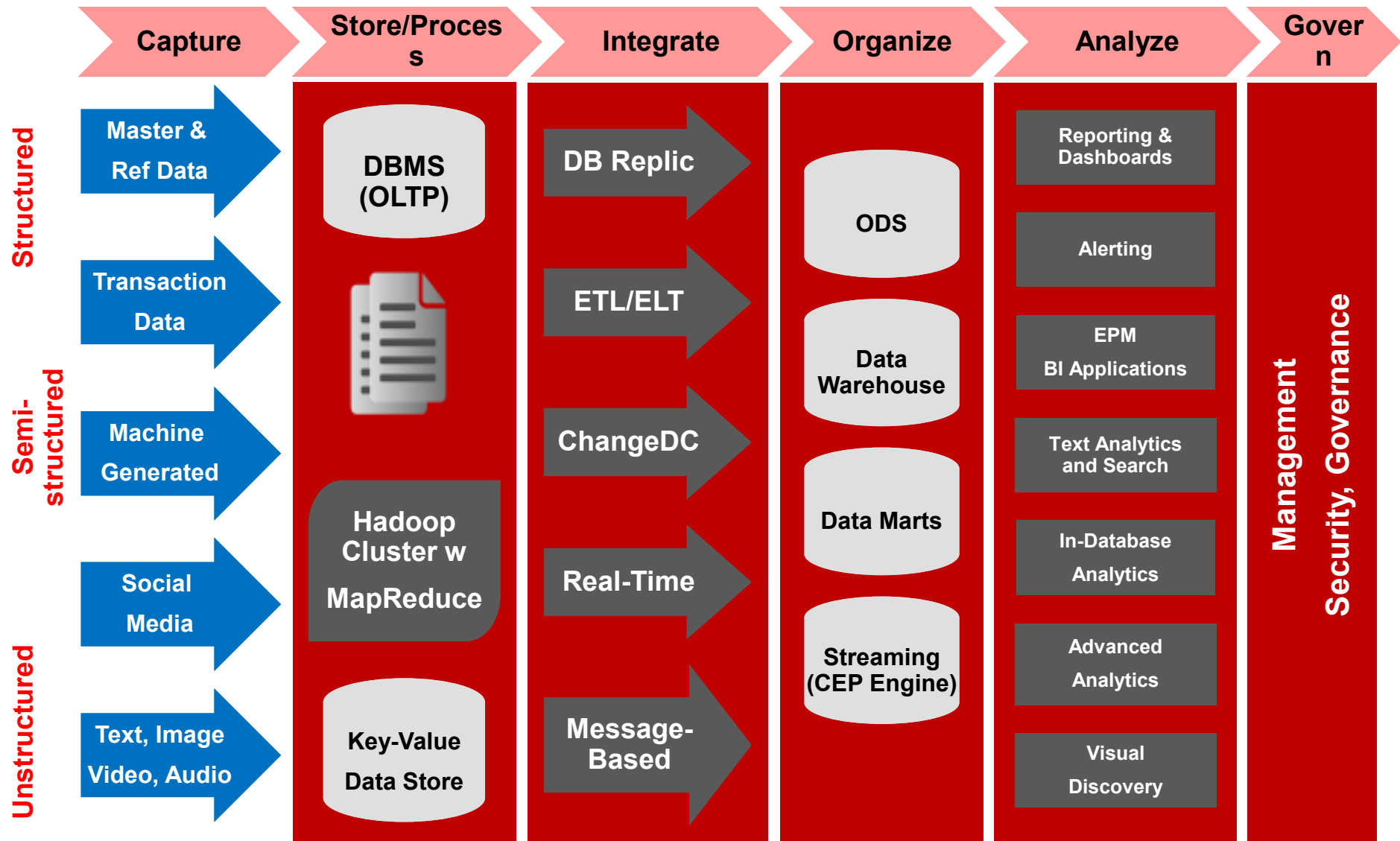
Batch Usage Model





SUMMARY

Big Data, An Integrated Architecture



ORACLE®