

ORACLE®

Big Data and Analytics in Government

Nov 29, 2012

Mark Johnson

Director, Engineered Systems Program



Agenda

- What Big Data Is
- Government Big Data Use Cases
- Building a Complete Information Solution
- Conclusion

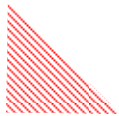
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.

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

Big Data Summary

Big Data Defined

“ Big data: techniques and technologies that enable organizations to effectively and economically analyze all of their data ”



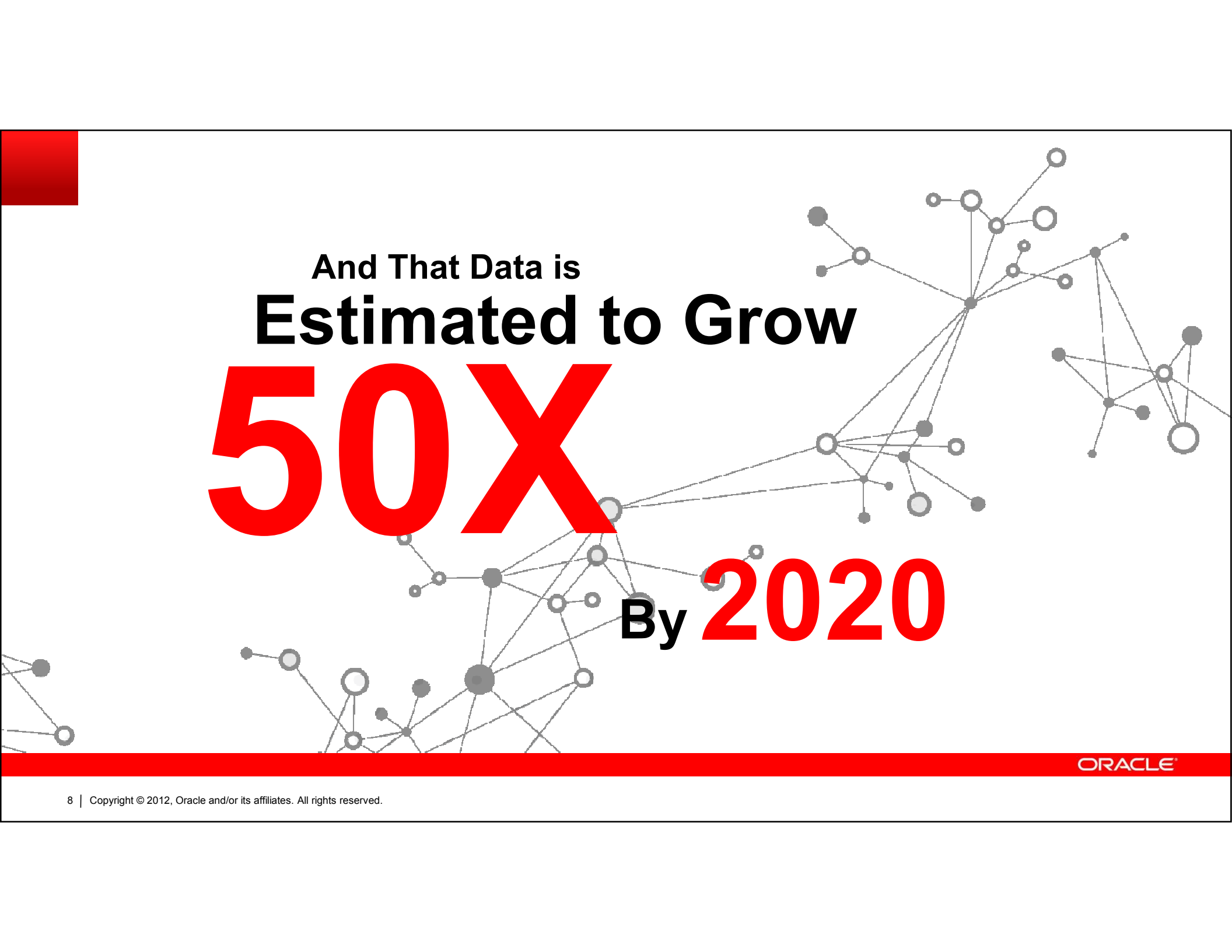


90%

**Of the World's Data
Has Been Created in the Last**

2 Years

ORACLE

A network diagram consisting of various sized grey circles (nodes) connected by thin grey lines (edges). The nodes are scattered across the slide, with a higher density in the upper right and lower left areas. A solid red square is located in the top-left corner of the slide.

And That Data is
Estimated to Grow
50X
By 2020

ORACLE

Social Data Generated Every Minute



695,000

Status updates



204,166,667

Emails



571

New
Websites

510,040

Comments



2,000,000

Search Queries



ORACLE

Big Data Buzz

“Why big data is a big deal”

InfoWorld – 9/1/11

“The challenge—and opportunity—of big data”

McKinsey Quarterly—5/11

“Ten reasons why Big Data will change the travel industry”

Tnooz -8/15/11

“Keeping Afloat in a Sea of 'Big Data”

ITBusinessEdge – 9/6/11

“Getting a Handle on Big Data with Hadoop”

Businessweek-9/7/11

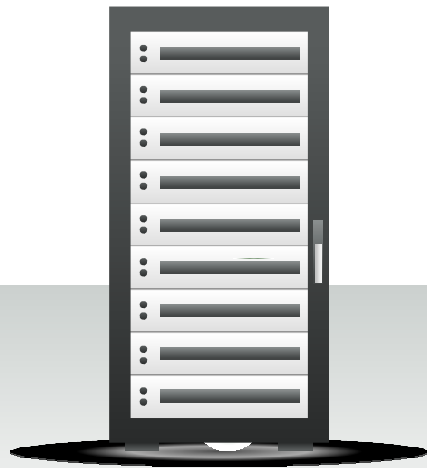
“The promise of Big Data”

Intelligent Utility-8/28/11

ORACLE

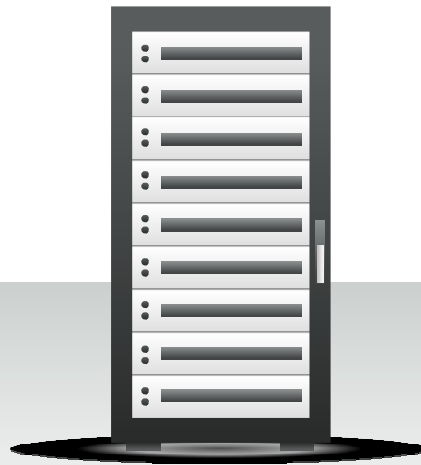
What is Big Data?

What Makes it Big Data?



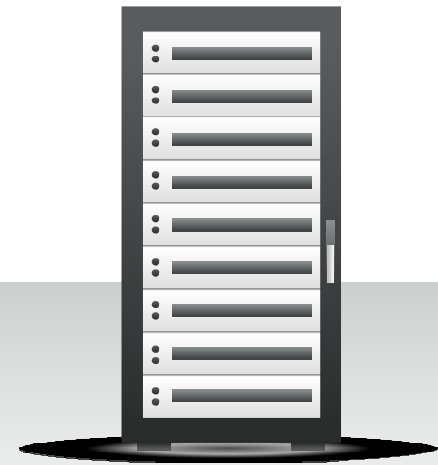
VOLUME

Very large quantities of data



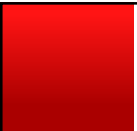
VELOCITY

Extremely fast streams of data



VARIETY

Wide range of datatype characteristics



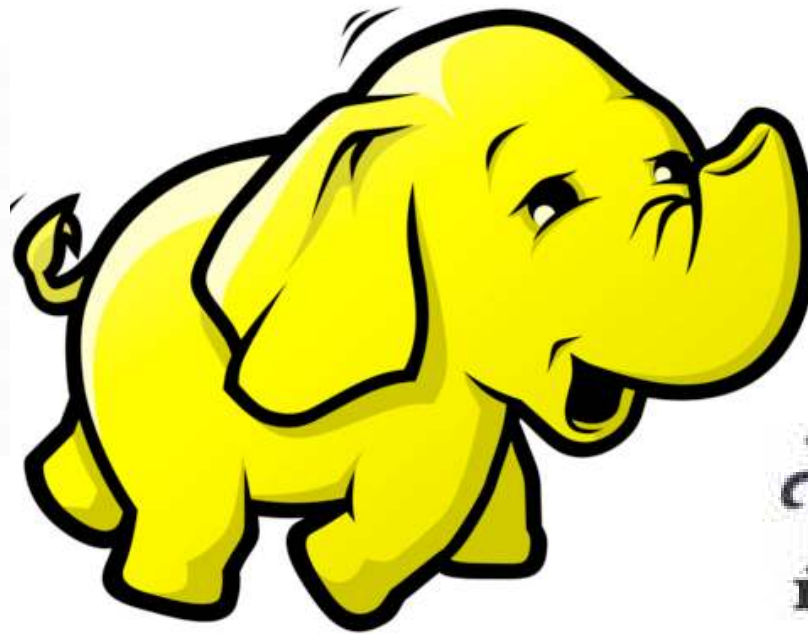
VALUE
IS HARD TO FIND

VOLUME

VARIETY



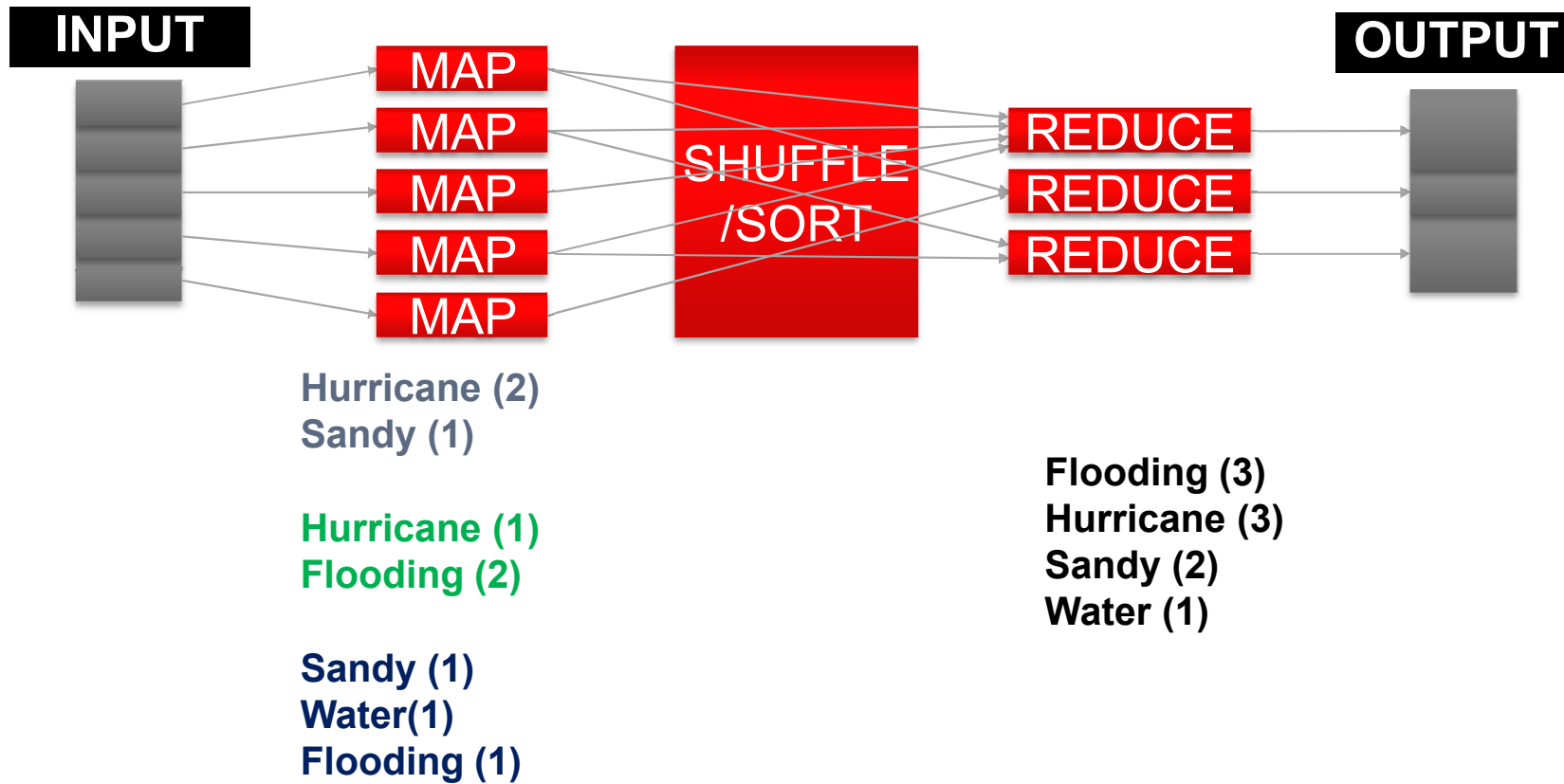
Big Data - Tools



Hadoop (Map/Reduce)

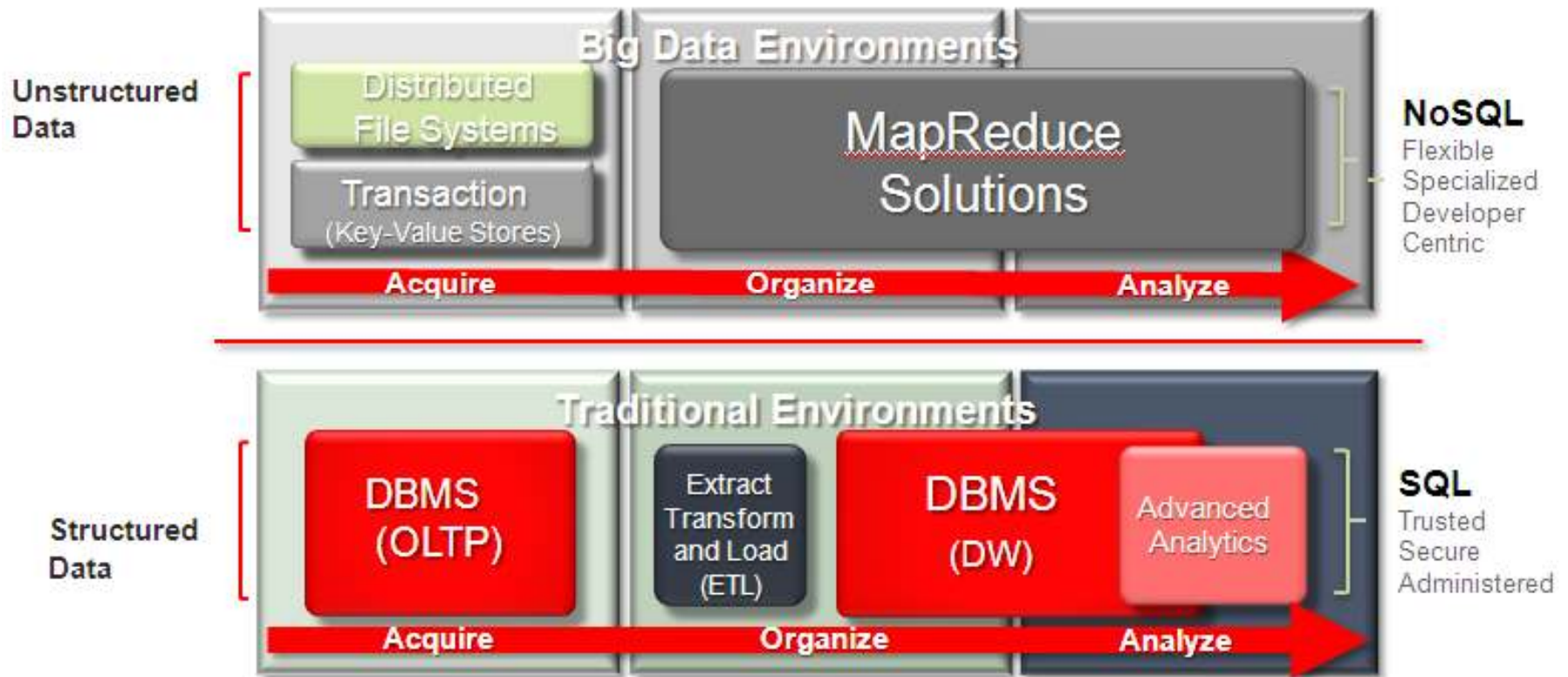


A Map/Reduce Pipeline



Big Data Tools Add to Existing Toolsets

New ways of processing data we couldn't access before



Government Big Data Use Cases

Big Data in Public Sector

Fraud Prevention



Revenue Management



Constituent Sentiment



Threat Identification



Economic Analysis



Healthcare



Regulatory Compliance, Licensing & Law Enforcement



Open Government



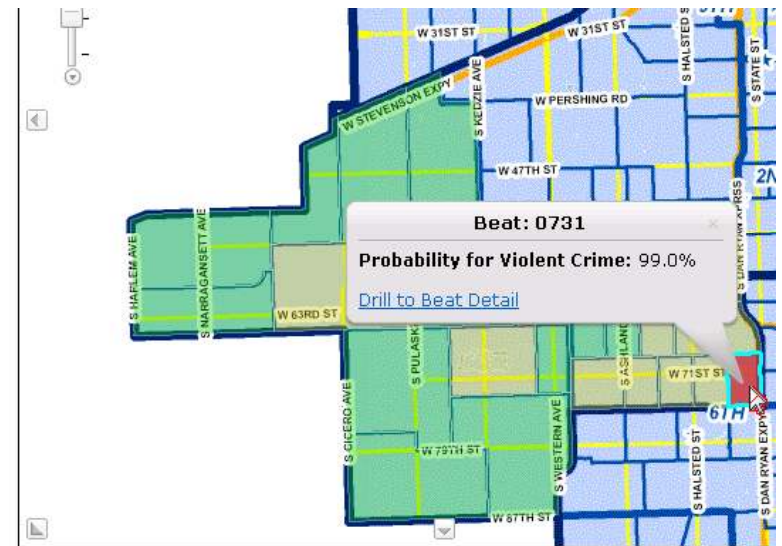
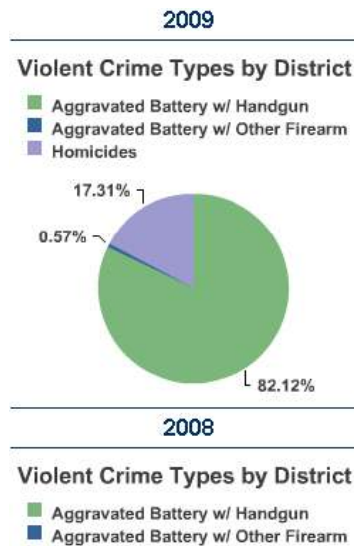
Maintenance & Utilities



Predictive Policing

Where is a violent crime likely to occur?

- Weather - recent and forecasted
 - Retrieved from the National Weather Service during nightly ETL processes
- Contact Cards
- 911 Calls
- Incidents
- Arrests
- Day of Week
- Date of Month



Big Data to detect Sales Tax Fraud

Potential revenue losses approach \$2.8B/year in one state

Supply Data



Credit Info



Tax Databases

County	Pct Change 08-09	Total \$200k plus in 2009	Total \$200k plus in 2008	\$200k-\$500k	\$500k-\$1M	\$1million +
Manhattan	-11.7%	81,455	92,267	54,252	14,990	12,415
Westchester	-7.6%	43,229	46,766	30,275	7,800	5,145
Nassau	-6.7%	43,876	47,029	33,666	6,471	3,719
Suffolk	-6.5%	29,125	31,134	23,941	3,429	1,755
Kings	-6.5%	17,471	18,644	14,204	2,078	1,189
Erie	-5.5%	7,670	8,119	6,110	1,075	485
Queens	-11.2%	11,074	12,477	9,807	905	362
Monroe	-8.3%	7,167	7,904	5,993	825	349
Rockland	-6.6%	7,309	7,822	6,141	789	379
Orondaga	-4%	4,539	4,727	3,729	562	248
Richmond	-10.4%	4,880	5,444	4,209	454	207
Albany	-3.2%	3,385	3,497	2,744	444	197

Labor Costs



Big Data for Building Audits

Targeting over-occupied buildings

Neighborhood Socioeconomic Status



Foreclosure Proceedings

Year of Construction



Tax Delinquency

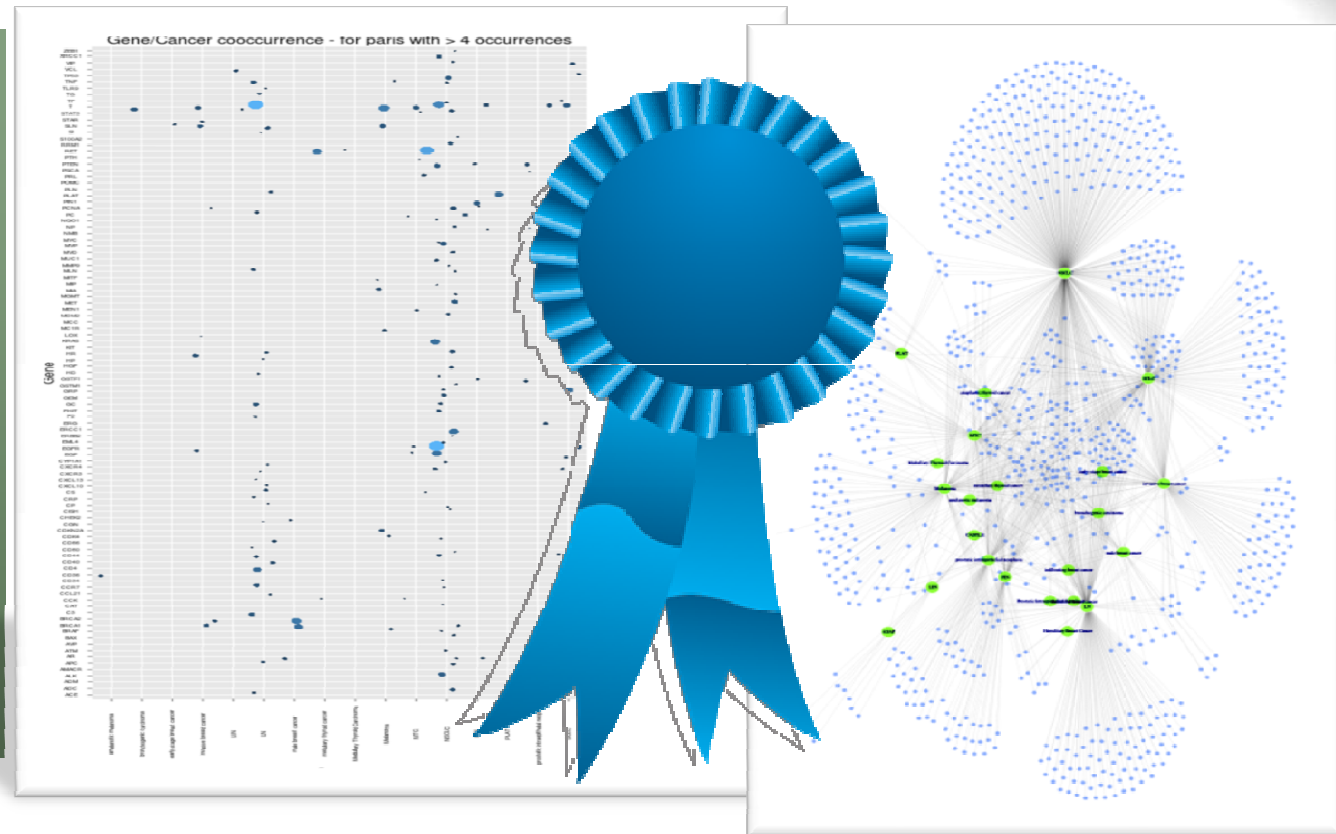
County	Pct Change 08-09	Total \$200k plus in 2009	Total \$200k plus in 2008	\$200k-\$500k	\$500k-\$1M	\$1million +
Manhattan	-11.7%	81,455	92,267	54,050	14,990	12,415
Westchester	-7.6%	43,229	46,766	30,275	7,809	5,145
Nassau	-5.7%	43,876	47,029	33,686	6,471	3,719
Suffolk	-5.5%	29,125	31,134	23,941	3,429	1,755
Kings	-5.3%	17,471	18,644	14,204	2,078	1,189
Erie	-5.5%	7,670	8,119	6,110	1,075	485
Queens	-11.2%	11,074	12,477	9,807	905	362
Monroe	-9.3%	7,167	7,904	5,993	825	349
Rockland	-6.6%	7,309	7,822	6,141	789	379
Onondaga	-4%	4,539	4,727	3,729	562	248
Richmond	-10.4%	4,880	5,444	4,209	454	207
Albany	-3.2%	3,385	3,497	2,744	444	197

New York City improved audit rates from 13% to over 70%

Big Data in Healthcare

Find relationship between gene to cancer interaction

- Cross-referenced the relationships between 17000 genes and five major cancer types across 20 million medical publication abstracts
- Cross-referenced genes from 60Million patients and miRNA for a simulated 900Million population.
- Understanding additional layers of the pathways these genes operate in and the drugs that target them is expected to help researchers in their work



Policy Implications of Big Data in Government

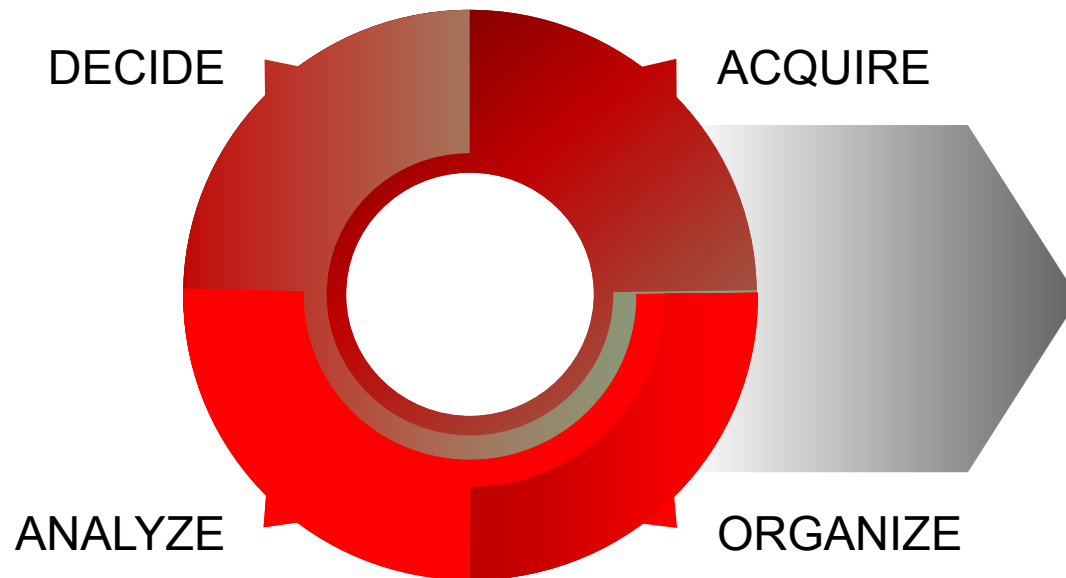
- What information can/should the government collect & aggregate?
- How is that information (and the resultant data) protected?
- How are errors identified & corrected?
- How is data collected by private companies accessed?
- Which government agencies can use the data?
- Etc...



Building a Unified Information Management Solution

ORACLE

BIG DATA LIFECYCLE



NEW REQUIREMENTS

New Tools for Acquiring & Organizing Information

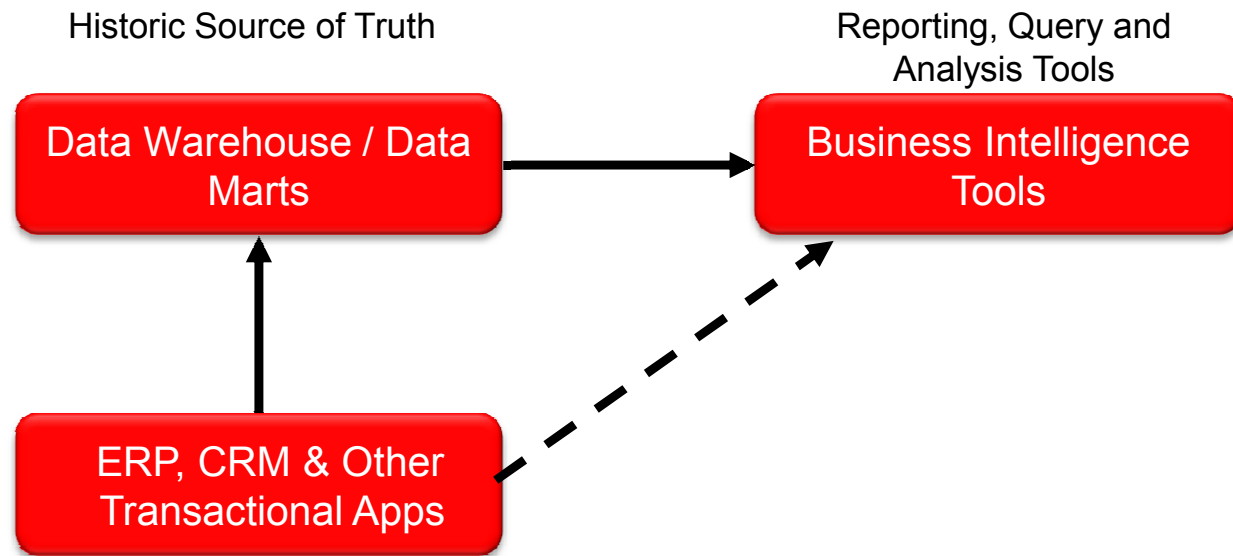
Easy Integration of New Infrastructure

In-memory Processing

Advanced Analytics

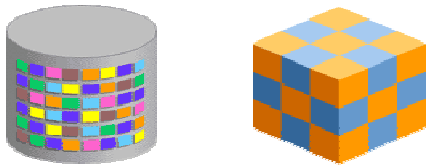
ORACLE

Footprint in Most Organizations Today



Goal: Make an Informed Recommendation

Structured Sources



Unstructured Sources



Constituent Analytics

Constituent History

Job Analytics

Position Inventory



Constituent Behavior

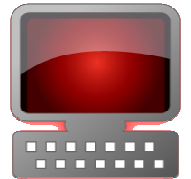
Sentiment & Influence

Channel Impact

Job Placement

Real-Time
Recommendations

Analytics
Search
Promotions



ORACLE

Using all Data to Understand Constituents

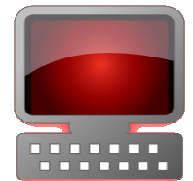
Data Warehouse



Website Logs & Data

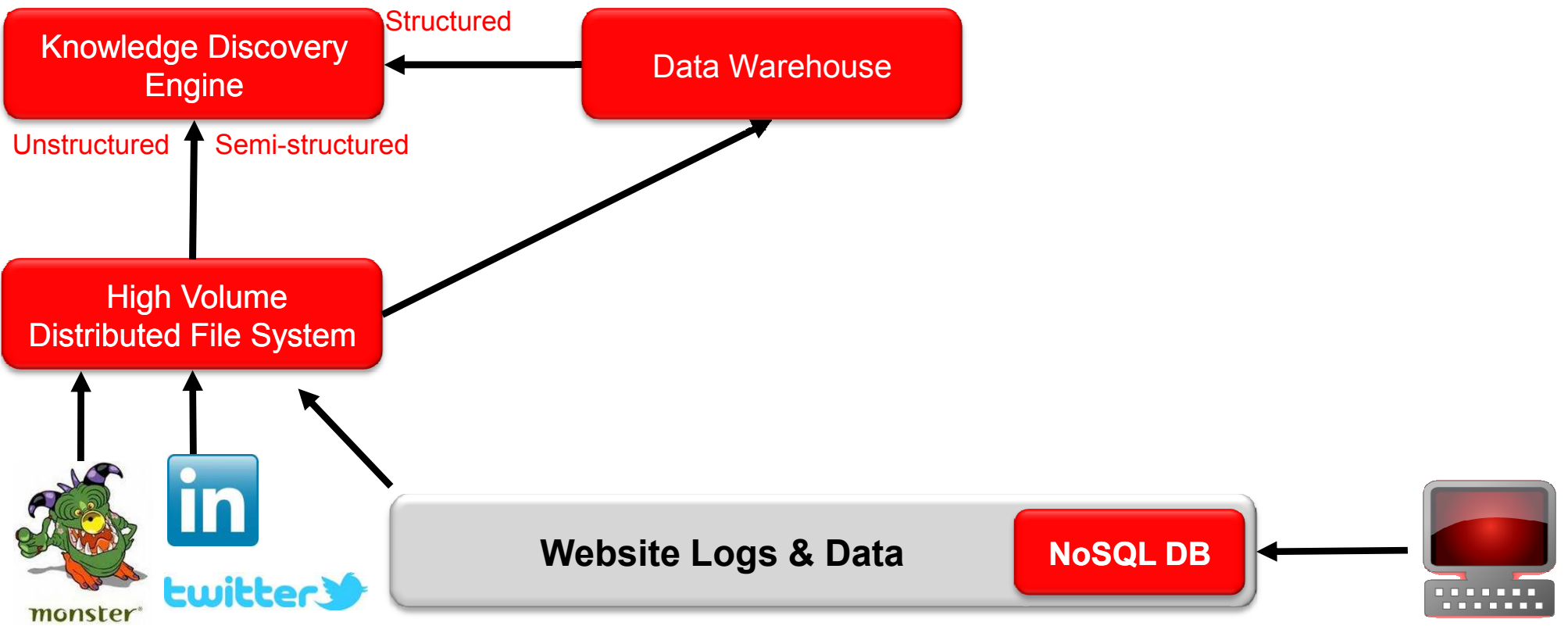
NoSQL DB

Employment Site



ORACLE

Discovering Valuable Data



Information Discovery Engine

Overview

The screenshot displays the Oracle Endeca Information Discovery interface. The top navigation bar includes the Oracle logo, the product name 'Endeca Information Discovery', and a 'Welcome!' message. Below the navigation bar, the interface is divided into several sections:

- Search Section (Red Box):** Contains a search box with the text 'Search', a search icon, and a 'Search Within' checkbox. A red box highlights this area with the word 'Search' overlaid.
- Range Filters Section (Green Box):** Contains a list of filters under 'Range Filters' and 'Guided Navigation'. The 'Guided Navigation' section is expanded to show 'Social' filters. A green box highlights this area with the text 'Guided Navigation' overlaid.
- Summary Statistics Section (Blue Box):** Displays key metrics: 'Total Interactions' (362,002), '+VE Sentiment' (46,742), '-VE Sentiment' (106,217), and 'Neutral Sentiment' (155,267). A blue box highlights this area with the word 'Analytics' overlaid.
- Visual Analysis Section (Blue Box):** Contains two bar charts: 'Total Interactions by content entities themes' and 'Total Interactions by author username'. Both charts show the top 10 items by total interaction. A blue box highlights this area with the word 'Analytics' overlaid.

The 'Total Interactions by content entities themes' chart shows the following data:

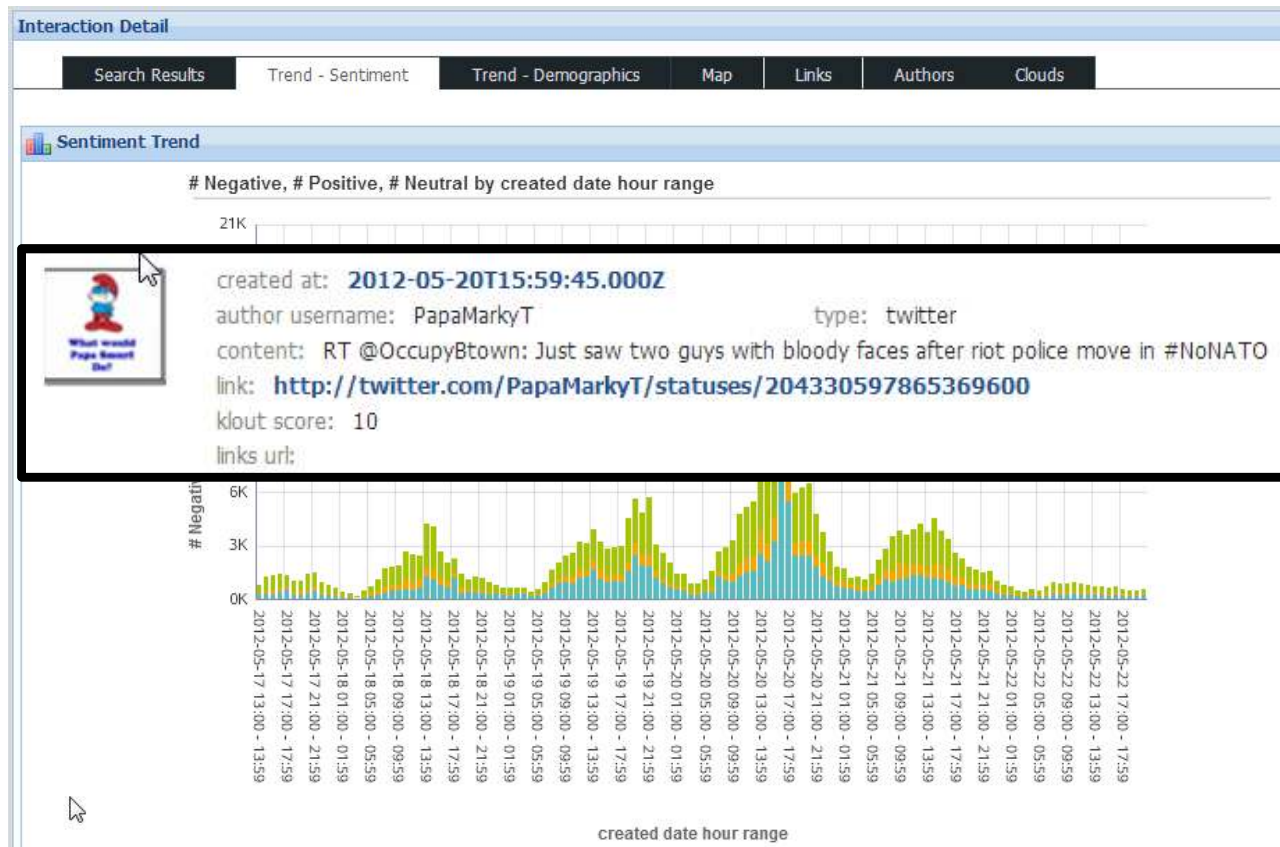
Content Entities Themes	Total Interactions
summit http	1200
police state	800
sound cannon	700
world leaders	600
supply routes	500
final day	400
terror plot	300
terrorism charges	200
afghan withdrawal	100
afghan war	100

The 'Total Interactions by author username' chart shows the following data:

Author Username	Total Interactions
OccupychicagoRT	1100
rightnowon_linead	900
sdcjaw	800
Occupycfics	600
diffipolis	500
Occupychicago	400
extenflyville	300
Kayme	200
dienertedman	100
youpositivetrk	100

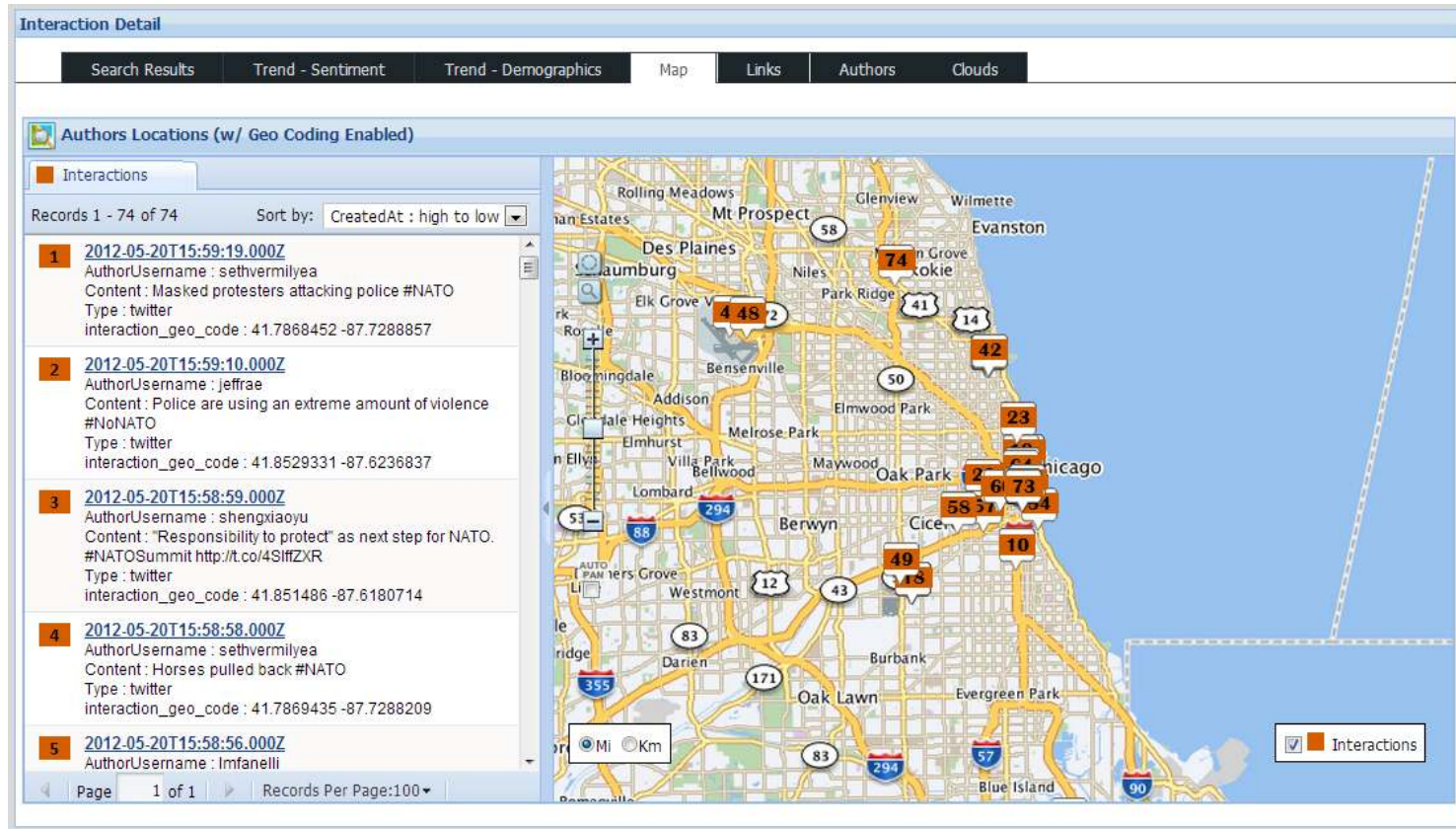
Information Discovery

Case Study

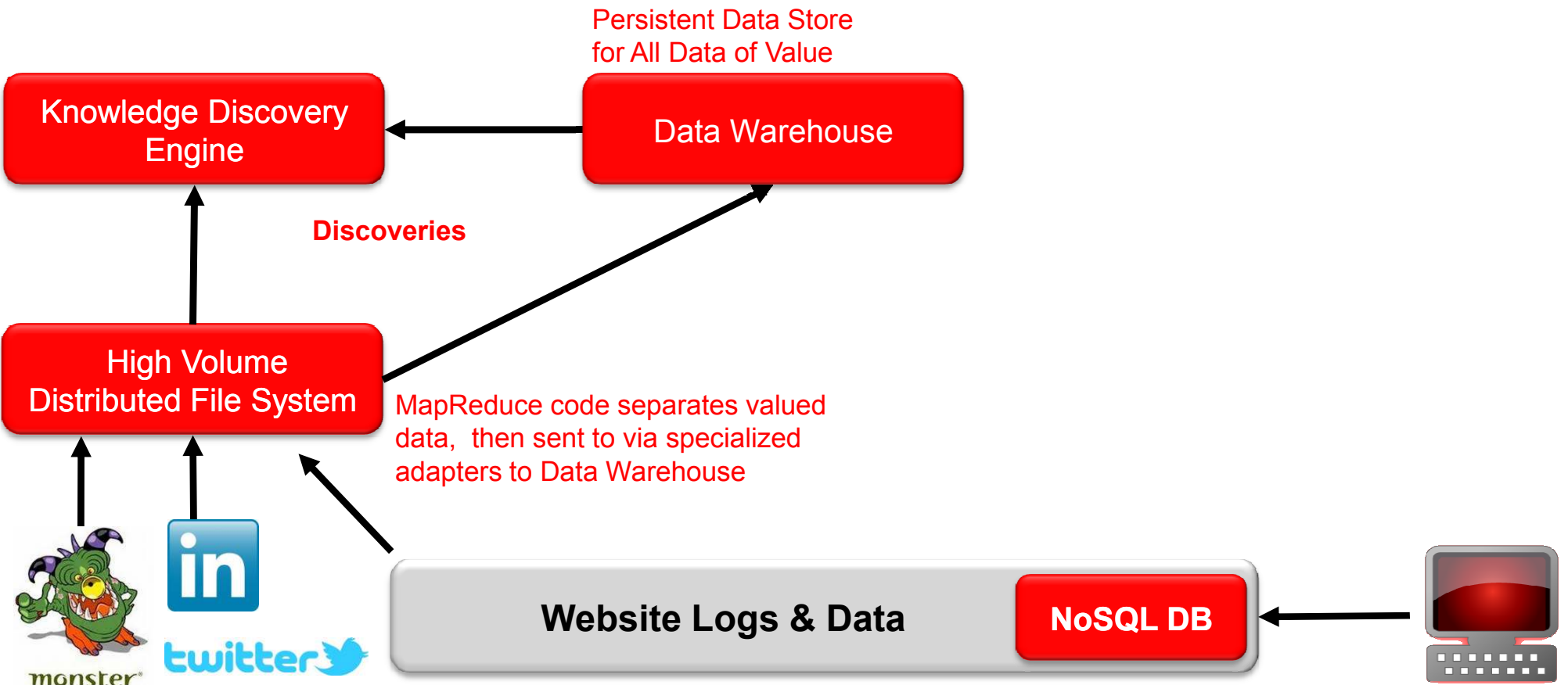


Information Discovery

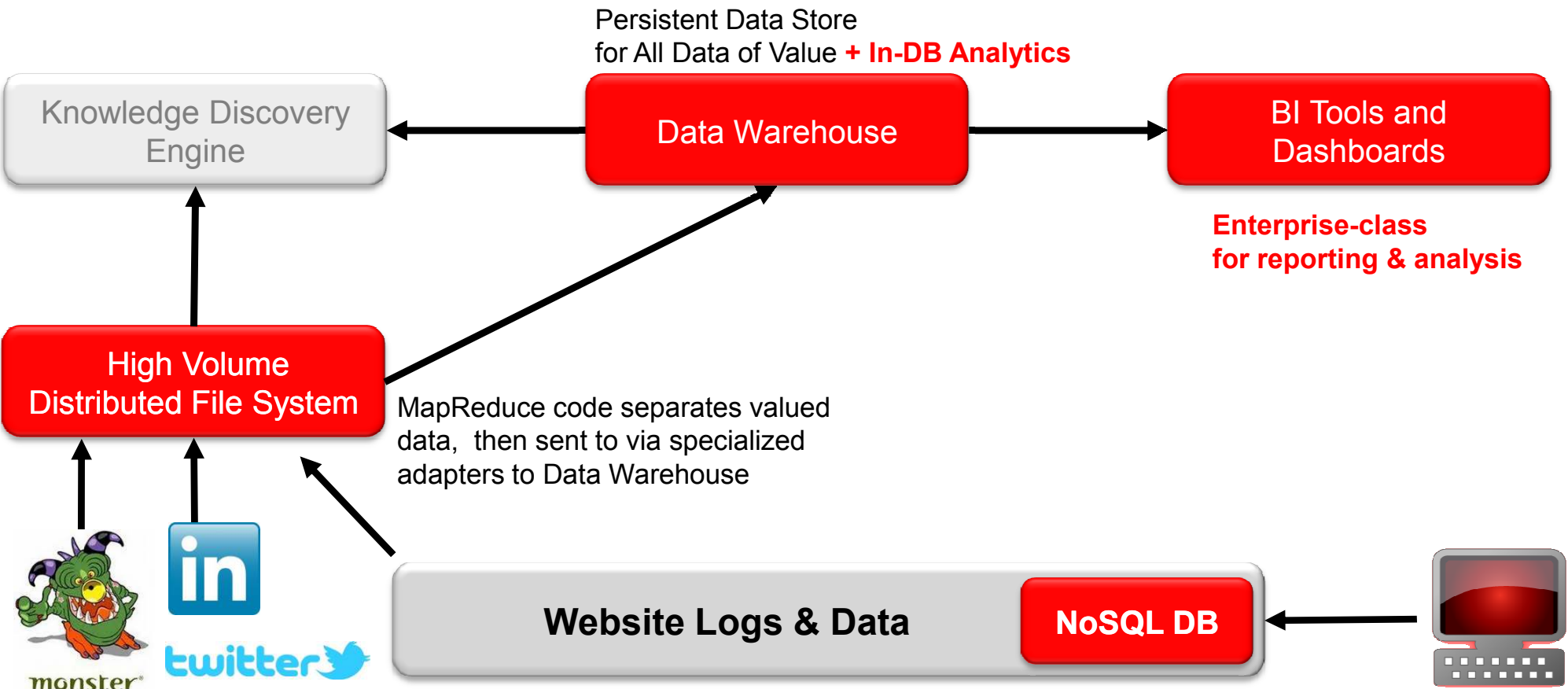
Case Study



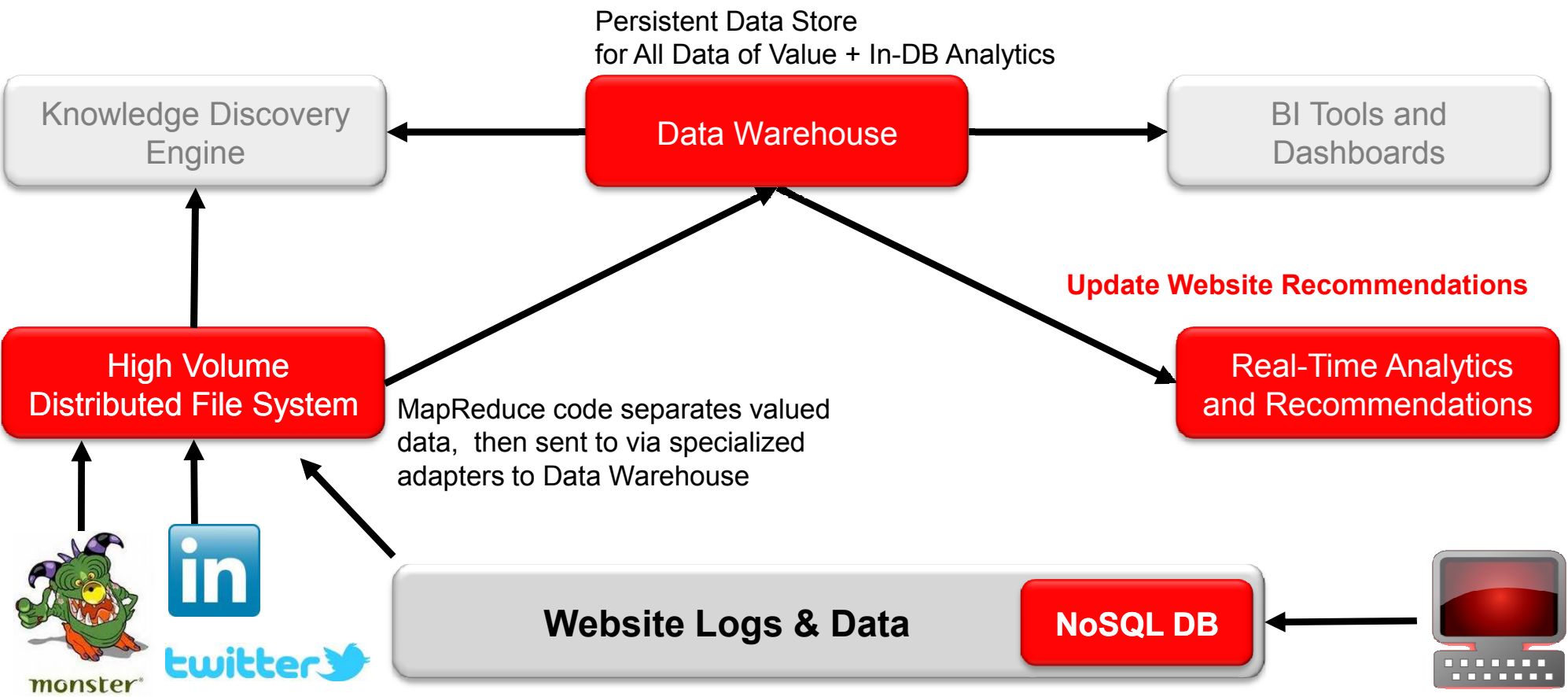
Valuable Data Found – Now Store it Securely



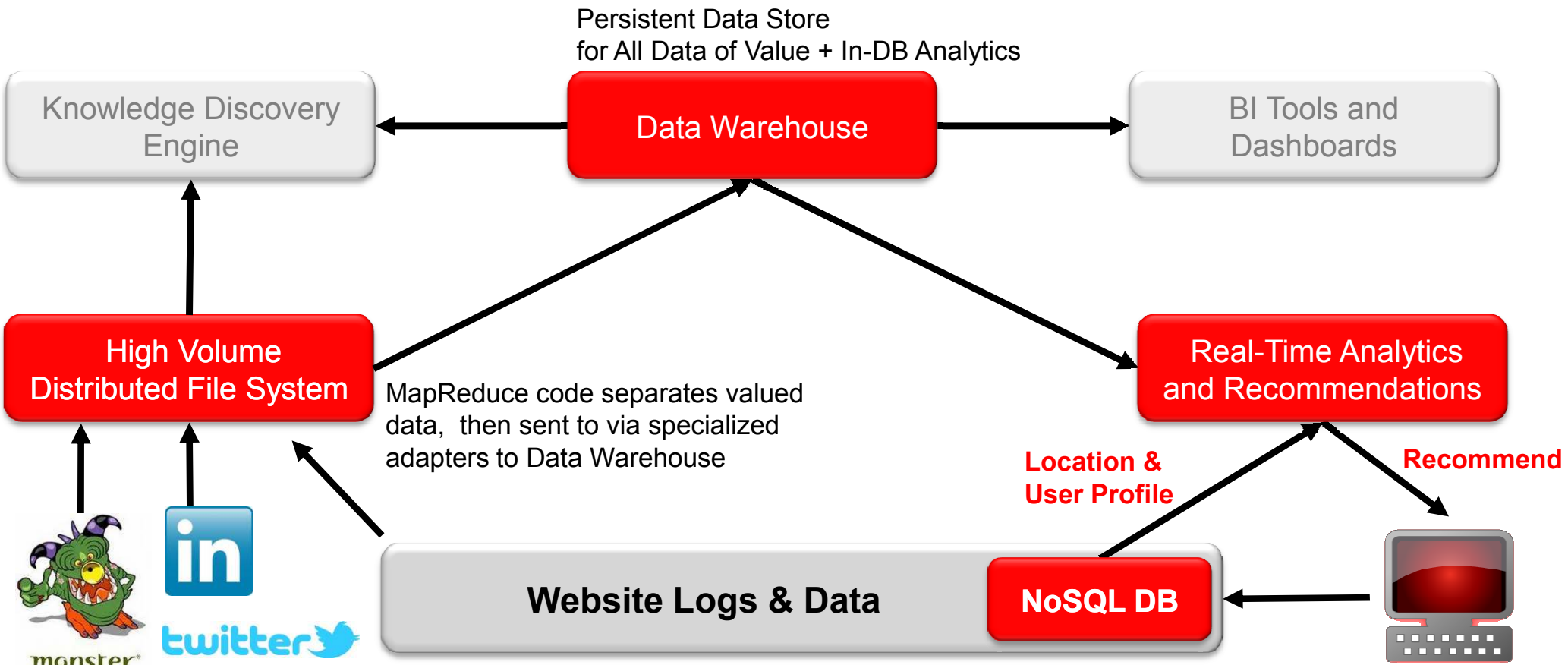
Deploy Widely Available Reports & Analytics



Feed the Recommendation Engine



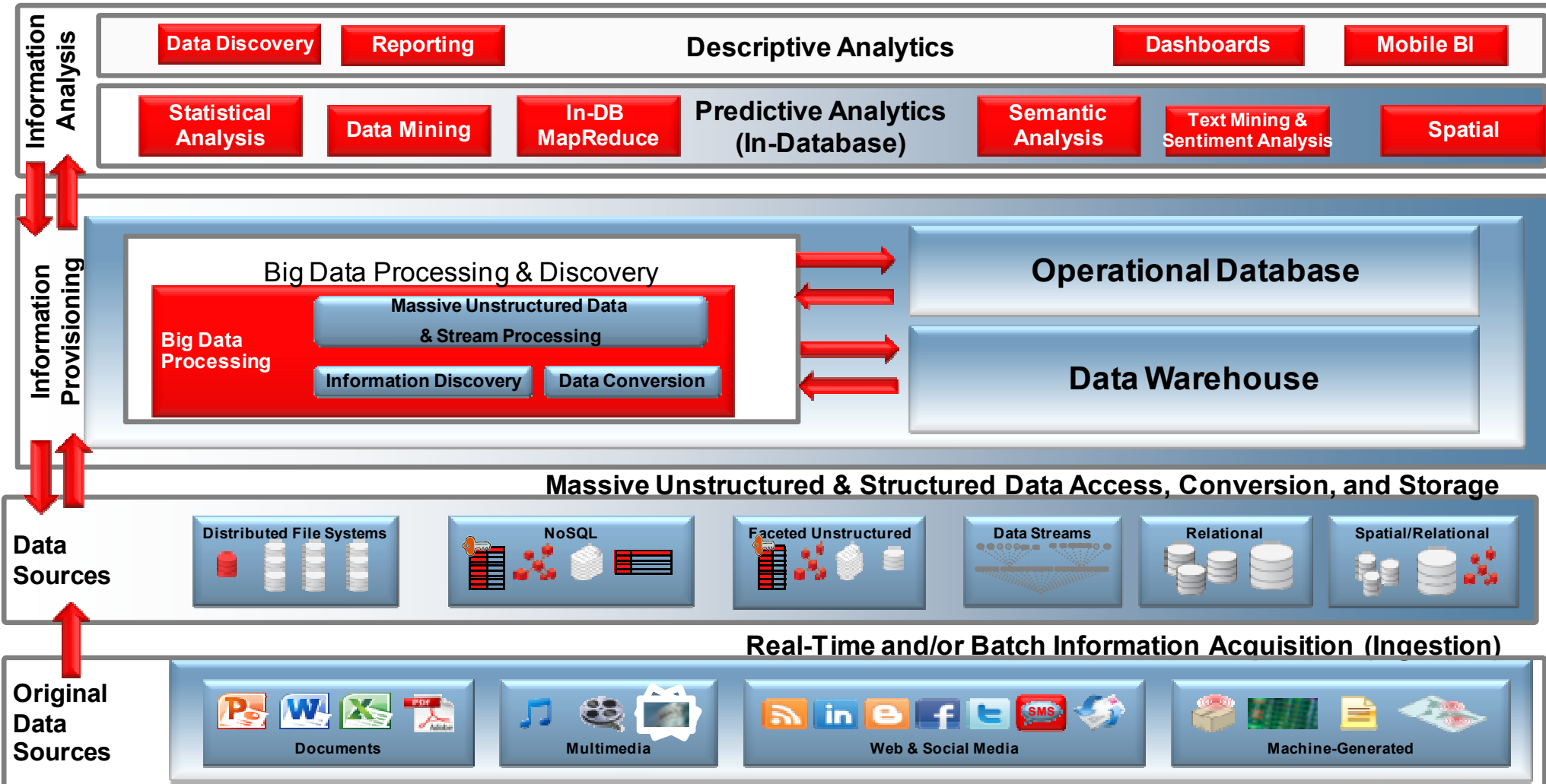
Make Well-Tuned Real-Time Recommendations



ORACLE

Summary

Comprehensive Data Reference Architecture



How do I Start?



- Think Big and Start Small
 - Find a question or requirement that your organization has been wrestling with responding to
- Establish a clear scope
 - Try for a quick win
- Use tools to flatten the initial learning curve
 - Setting up a Hadoop cluster is a specialized skill set
- Scale up as necessary



Mark.A.Johnson@Oracle.com

ORACLE

Hardware and Software

ORACLE®

Engineered to Work Together

ORACLE®