

ORACLE®

# New Oracle Business Analytics technologies for new challenges

Ljiljana Perica, Oracle Business Solution Leader

[Ljiljana.Perica@oracle.com](mailto:Ljiljana.Perica@oracle.com)

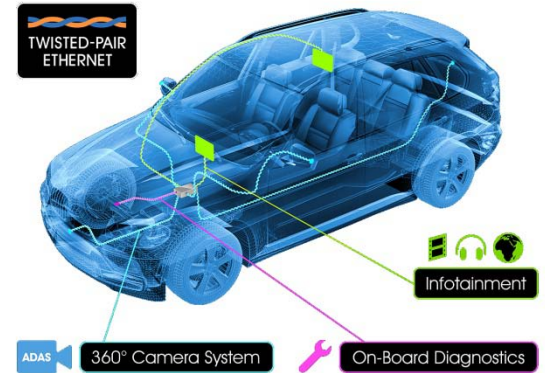
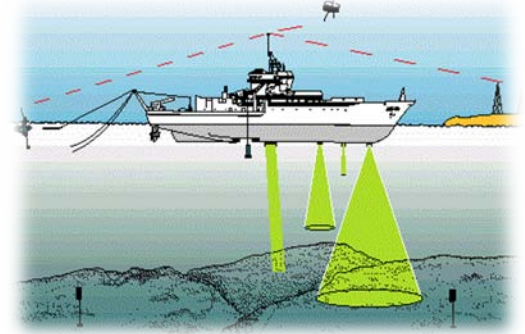


CONVERGE.  
CONNECT.

Empower the Modern Enterprise

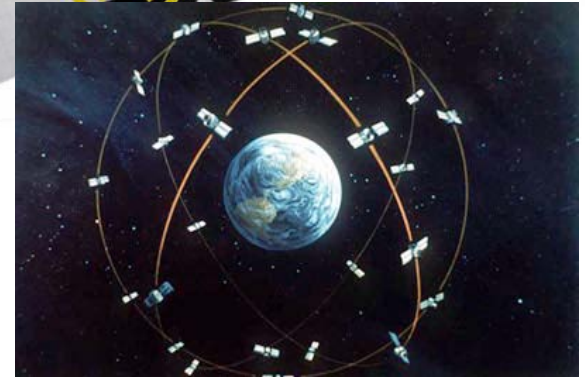
# Where is the **New** Information?

## *Born Digital: Sensors and Clickstream Logs*



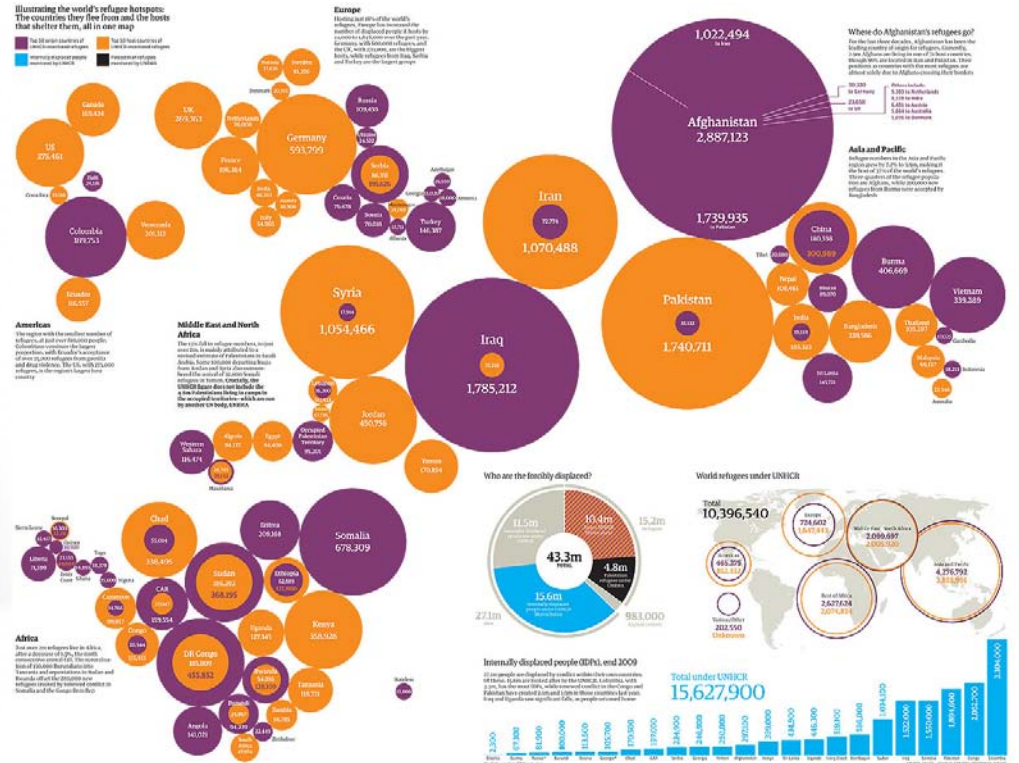
# Where is the **New** Information?

*Born Digital:* **Social Media, Blogs, M2M, GPS**



# Where is the **New** Information?

## *Born Analog: Open Data, Data without metadata*



# World of Big Data

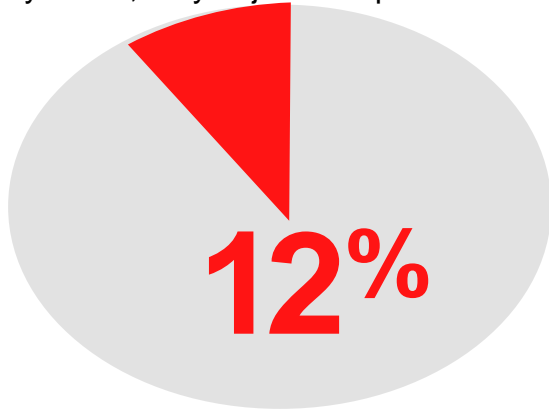


“The real **innovation** here is that we can **ask questions** and get the **answer back** **before** we have forgotten **why** we asked the question in the **first place.**”

– **Hilary Mason**, Chief Scientist **Bit.ly**  
+ member of NYC Mayor Bloomberg’s Technology and Innovation Advisory Council

# The Problem

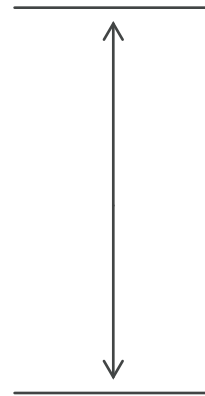
- While we are creating and collecting mountains of data, our ability to produce it has outstripped our ability to use it
- In many businesses: the information they need to improve products and services already exists, they're just not quite sure how to use it



Executives who feel they understand the impact data will have on their organizations

*Study done by The Economist Intelligence Unit and Oracle*

**Produce  
Data**



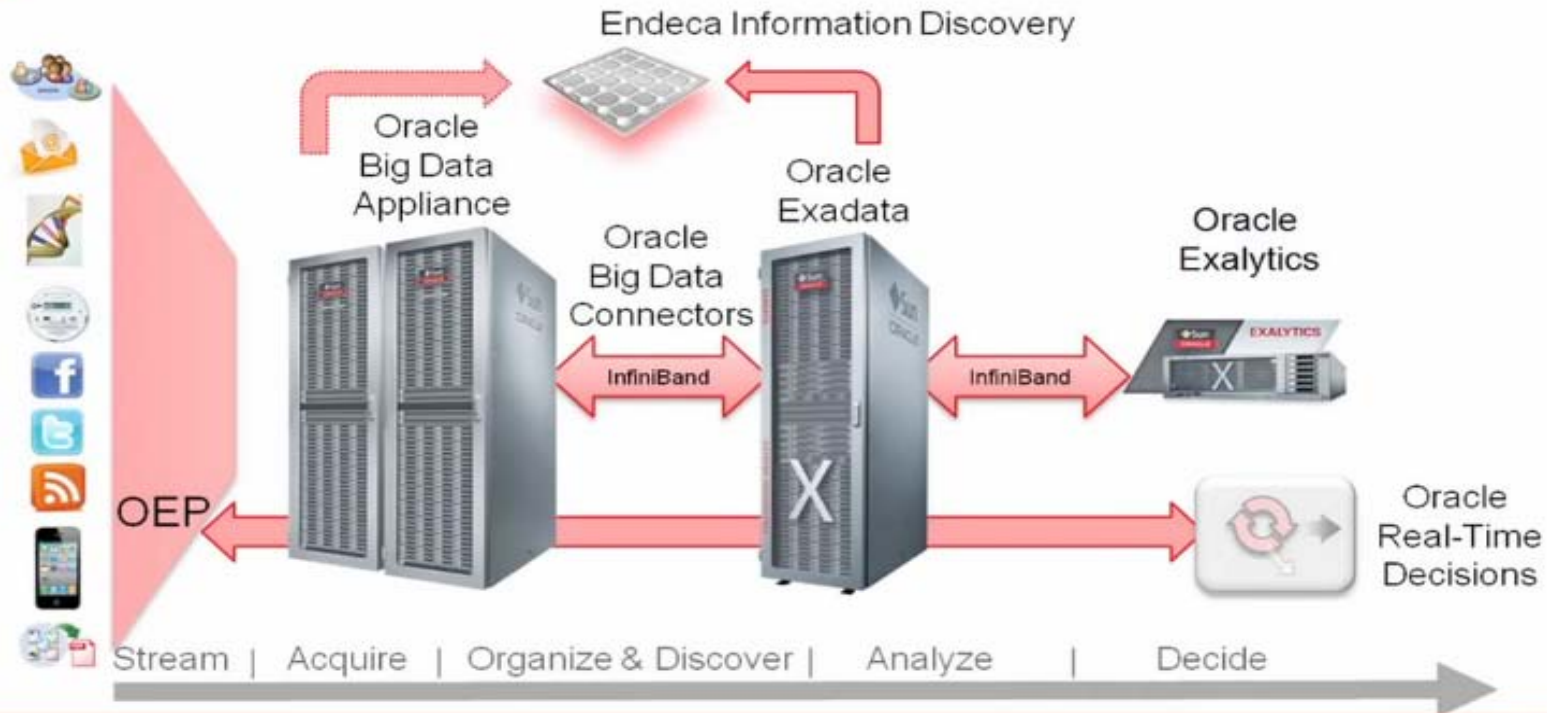
**Use  
Data**

“...The more companies characterize themselves as data-driven, the better they performed on objective measures of financial and operational results. In particular, companies in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors.”

### **Big Data: The Management Revolution**

By Andrew McAfee and Erik Brynjolfsson (MIT)  
Harvard Business Review, October 2012

# Oracle's Big Data Solution



ORACLE

ORACLE



“Unstructured content represents as much as 80% percent of an organizations total information assets. While Big Data technologies and techniques are well suited to exploring unstructured information, this ‘Big Content’ remains grossly underutilized and its potential largely unexplored.”

## **Gartner**

Darin Stewart (Research Director) | May 1, 2013

Zdroj: <http://blogs.gartner.com/darin-stewart/2013/05/01/big-content-the-unstructured-side-of-big-data>

# Unstructure Data Discovery

## Endeca Information Discovery

# Up to 80% of the data in organizations is unstructured

## The most important sources of unstructured data for business analysis

*Unstructured data has no pre-defined data model and/or does not fit well into relational tables. Typically, there is no identifiable structure – it can have complex, hierarchical structures, and is often text-heavy*



**Call Center/ Service Desk App.**  
(notes, comments, records,  
insurance claims, activity logs)



**Medical Records**  
(patient card, records, notes,  
documentation)



**Facebook**  
(discussion, comments,  
posts, status, Klout score)



**Bank and card transactions**  
(user notes, generated  
descriptions)



**Sensors**  
(data and logs  
from sensors)



**RSS**  
(posts, comments)



**Log files**  
(web and application logs)



**Blogger™  
Blogs**  
(posts, comments)



**Web**  
(review, forum, posts,  
comments, discussion)



**Dokumenty**  
(internal/ external documents, notes,  
complaint, claims, CVs, requests,  
forms)

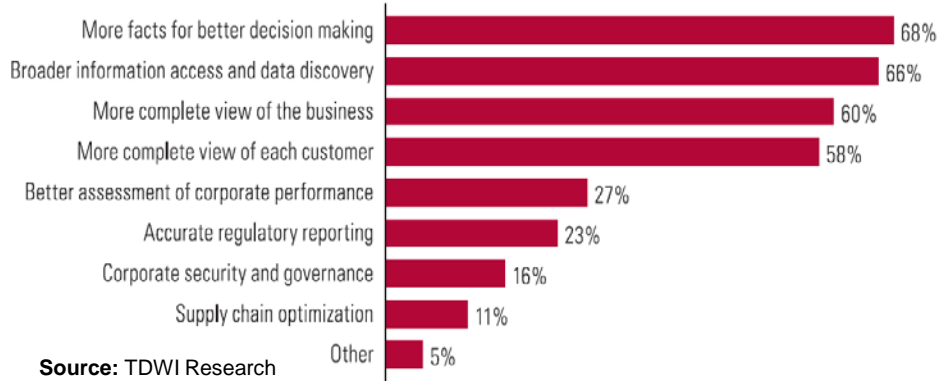


**E-mail**  
(internal / external  
communication)

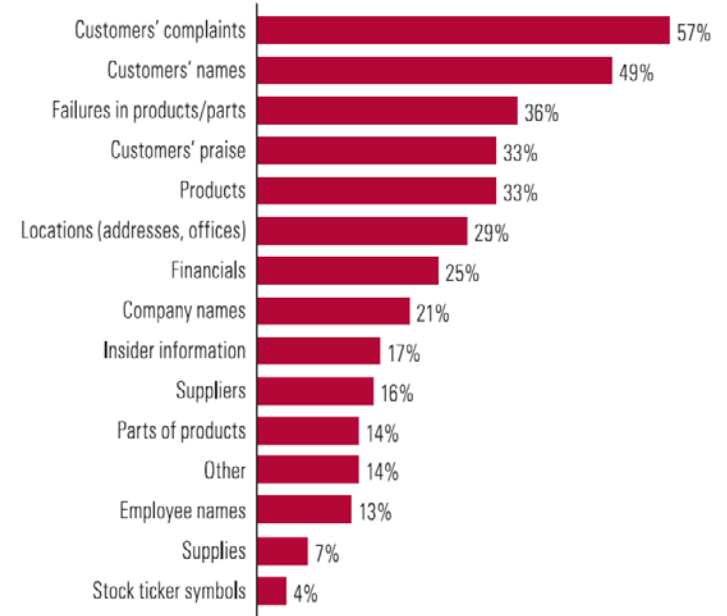
# Why think about unstructured data

## The main reasons from point of customers

How would BI users benefit from data extracted from text sources?



Which facts and entities are the most important for extraction from text sources?



FIRMS ABLE TO SUCCESSFULLY CAPTURE AND MANAGE THEIR UNSTRUCTURED DATA [hold a competitive advantage](#) over firms unable to do the same – over 90% of enterprises are currently planning it or are already doing it (DATAVERSITY 2012).

# Extend Business Analytics with Unstructured Data

## Introducing Oracle Endeca Information Discovery

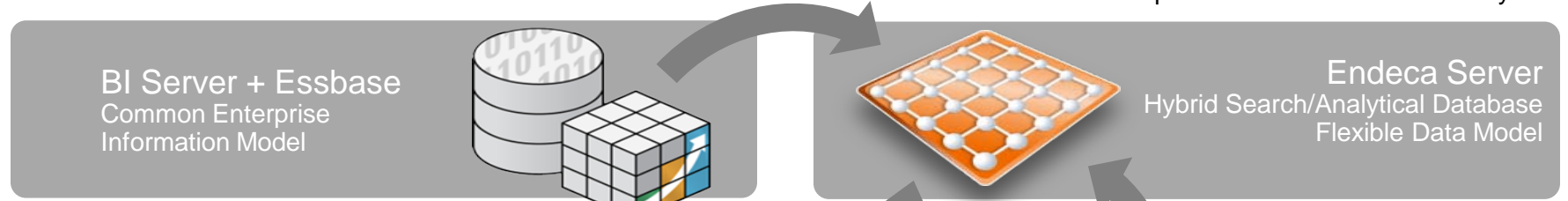
OEDI extends Oracle's analytic capabilities across unstructured sources, uniting the worlds of structured and unstructured to provide business users with complete visibility into their business processes, creating new insights and enabling better business decisions.

### Oracle Business Intelligence

Best platform for integrated ROLAP and MOLAP

### Oracle Endeca Information Discovery

Best platform for Unstructured Analytics



### Structured Data

OLTP & ODS Systems



Enterprise Applications (Oracle, SAP, Others)



Data Warehouse & Data Marts



### Unstructured Data

Websites



Content Systems, Files, Email



Social Media

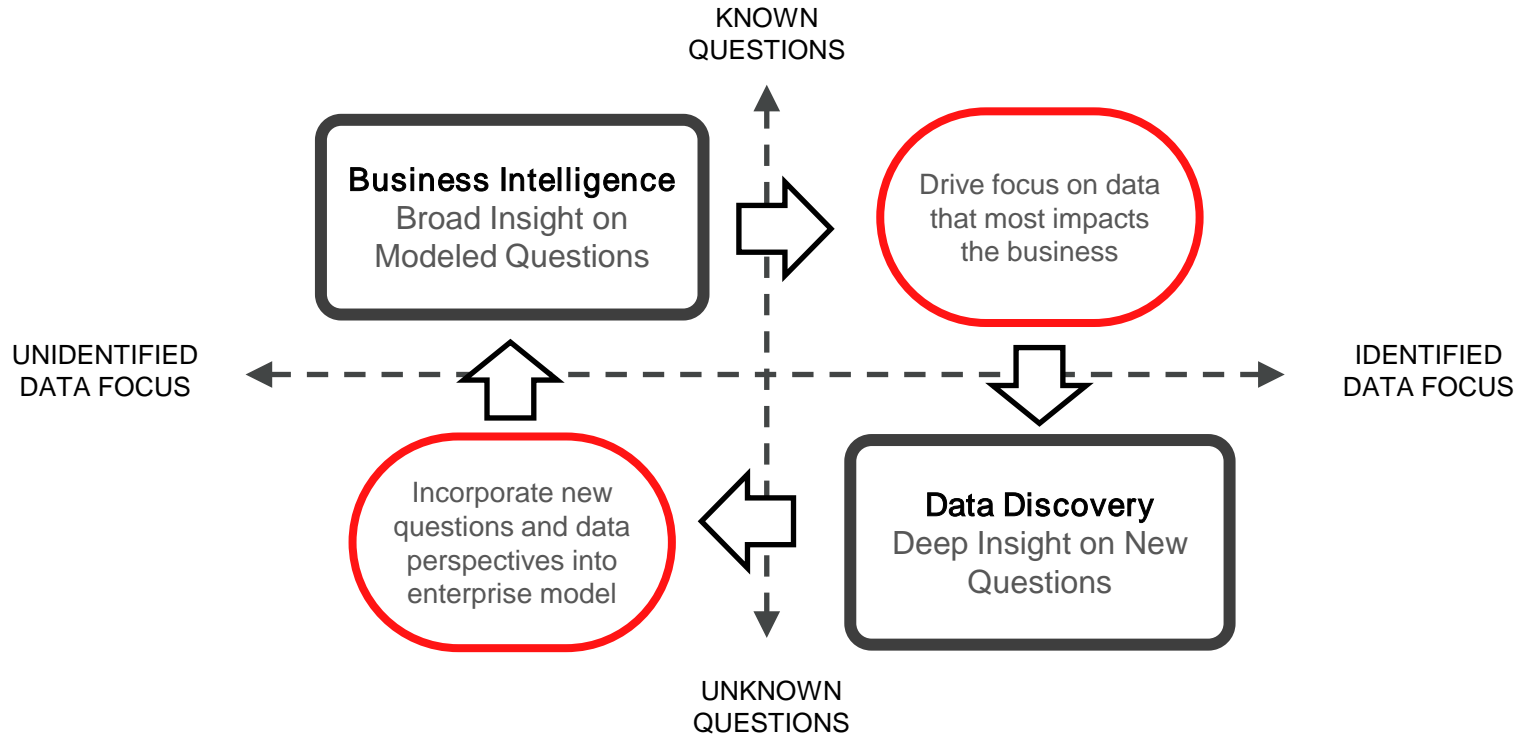


Big Data



# Business Intelligence and Data Discovery

## Distinct but Complementary Capabilities for Business Analytics



# Oracle Endeca Information Discovery

## Platform Technology Overview

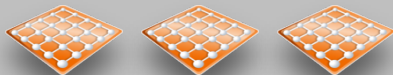
### Studio

Intuitive Exploration and Analysis  
Create and Share Apps



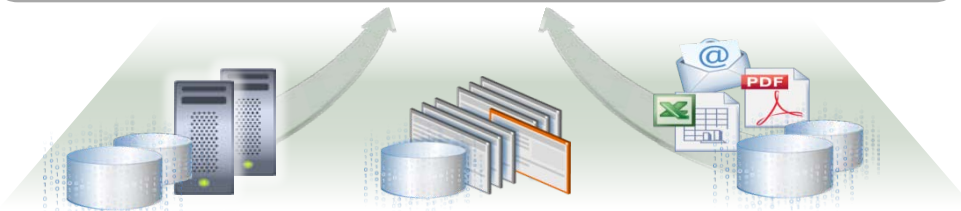
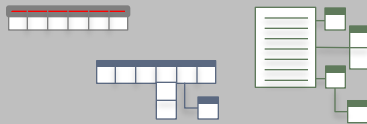
### Endeca Server

Hybrid Search/Analytical Database  
In-Memory Architecture



### Integration Suite

Data Integration and Enrichment  
Structured and Unstructured



### Studio – Web Application

- Contextual Search, Navigation, Analytics
- Qualitative and Outlier Visualizations
- Easy Drag-and-Drop Applications

### Endeca Server – Core Database

- Dynamic Data and Metadata
- In-Memory, Multi-Threaded Performance
- Enterprise Scale, Security
- Combining index-based engine & column-oriented in-memory database

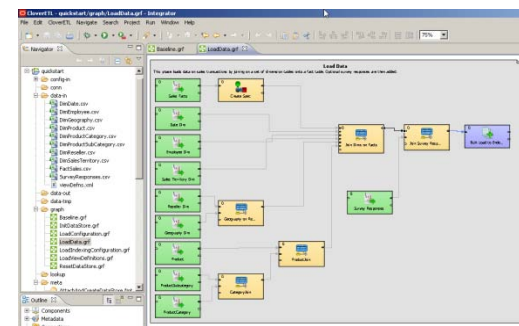
### Integration Suite – ETL

- Integrates Structured and Unstructured
- Text Enrichment and Sentiment Analysis

# Endeca Integration Suite

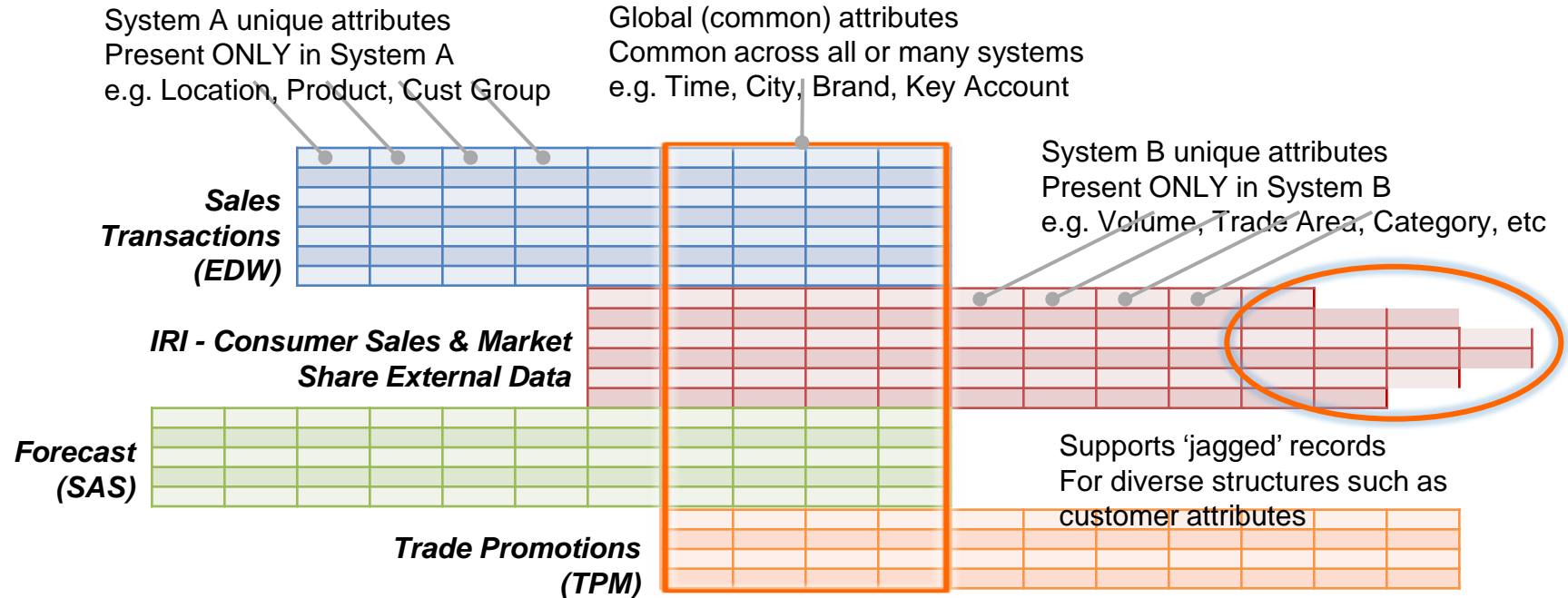
## Loading diverse information into Endeca Server

- **Integrator** - ETL environment, integrating and enriching enterprise data and content
- **Text Enrichment and Sentiment Analysis** - extract concepts, entities, and meaning from unstructured text fields in databases or other content systems for entirely new types of analysis
  - Text Enrichment = process of deriving structure from unstructured text, applying algorithms to extract entities, concepts, summaries, and sentiment, which are then appended to existing records as new fields. Oracle Endeca Information Discovery also provides basic term and regular expression whitelist tagging.
  - Sentiment Analysis is a subset of text enrichment, for deriving a sentiment score and metadata around to what degree the words represent positive, negative, or neutral emotion.
- **Integrator Acquisition System** - collection of connectors, integrating unstructured content from network file systems, Websites, and content management system repositories (EMC Documentum, EMC Documentum eRoom, FileNet P8, FileNet Document & Image Services, Interwoven TeamSite, Lotus Notes/Domino, Microsoft SharePoint, and OpenText LiveLink)
- **Open Interfaces and Connectors** – using Endeca Server’s open Web services API, allows direct data integration from industry-standard ETL tools, such as Oracle Data Integrator, Informatica PowerCenter, Apache Hadoop, and SOA-based services





# Endeca Server: Endeca brings any data together in a new way



- Employs a unique flexible data model that reduces the need for up-front modeling, enables the integration of diverse and changing data, and supports the broad exploration and analysis needs of business users
- Data model is based on the concept of "records"
- Does not employ an overarching schema for data - instead, every record has its own schema
- Support jagged data – each record can have its own set of unique attributes even if they are from the same source system
- End-users interacting with the Endeca application can use any starting point to enter the data, even a basic search term

# Endeca Studio: Simplicity of Search, Power of BI

## Interactive Data Exploration and Analysis

- Walk-up ease of use

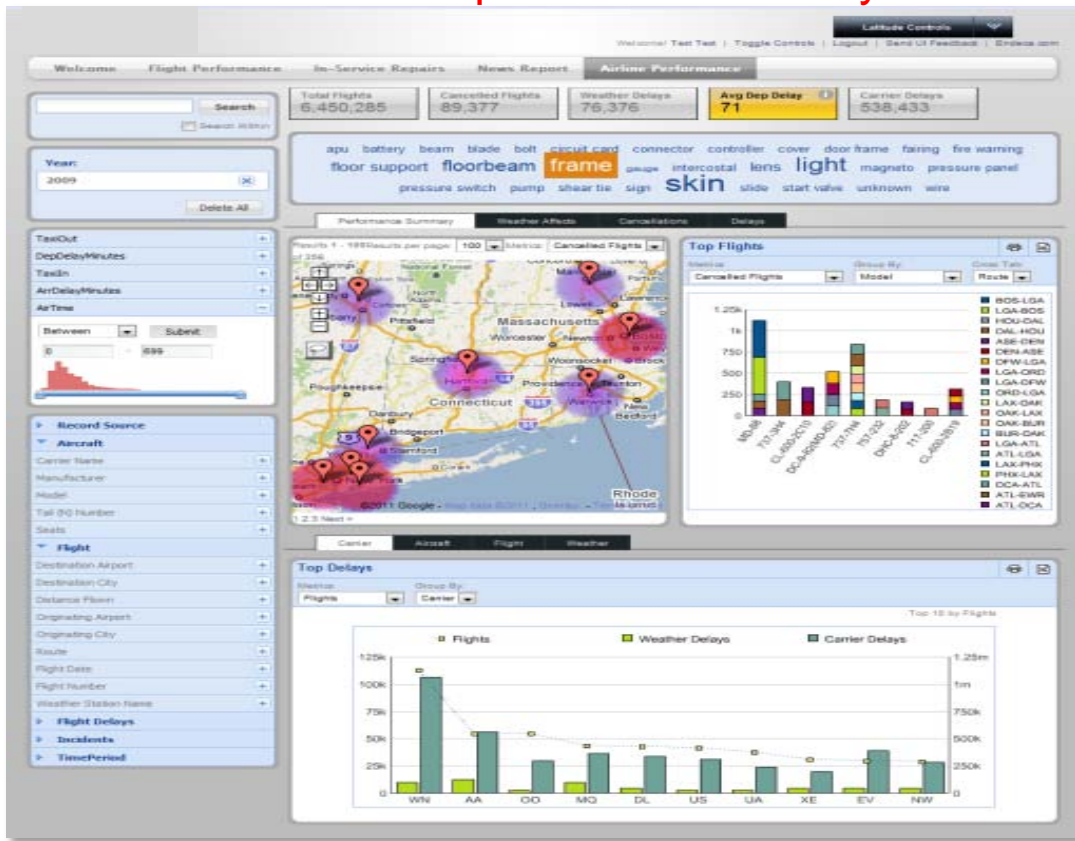
- Derived from 10 years of e-Commerce consumer user experience
- Familiar online experience

- Search + Faceted Navigation + Visual Analysis

- Search across *all* data and select attributes, like a web site
- Dynamic typeahead
- Automatic spell correction
- Unlocks unstructured data
- Data-Driven Contextual Navigation: Freely browse data without predefined paths or writing queries, Interactive: Shows only valid next steps
- Visual Analysis: Charts, crosstabs, key metrics, Geospatial visualization, Tag clouds

- Interactive Exploration

- Responsive Endeca Server



# Endeca Studio: Self-Service Data Mashup and Data Discovery

## Self-Service Data Mashup and Enhanced Discovery Dashboards

Personal Excel Files



Oracle BI Server Models



IT Provisioned Enterprise Data

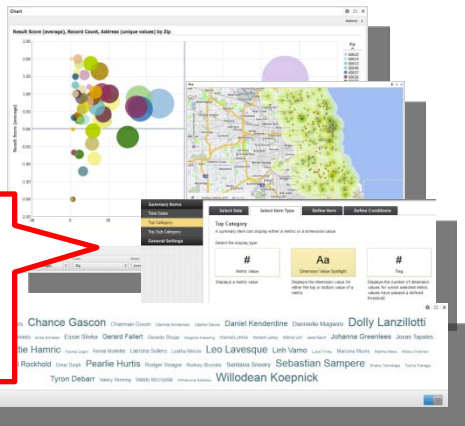


Cloud and Social

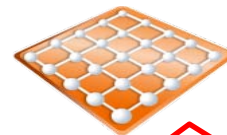


Mashup multiple sources including JDBC, Oracle BI Server, JSON, Hive

Intelligent defaults, flexible charts, heat maps, wizard driven components. Easier drag and drop and more...



## Enterprise-Class Data Discovery



ORACLE  
DATA INTEGRATOR

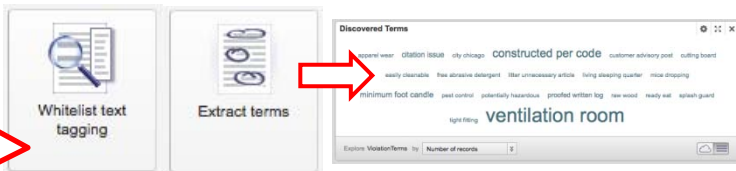


Faster performance, more scale, improved auditing, trusted data source library and increased connectivity

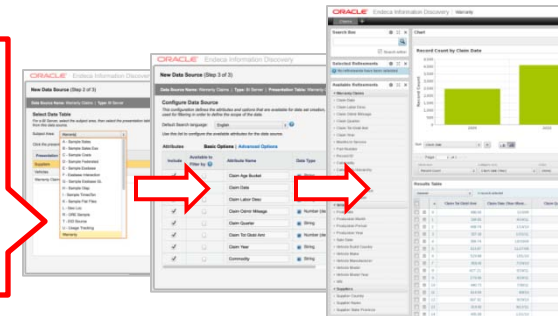
## Tighter Integration with Oracle BI

## Deeper Unstructured Analysis

End-users can enrich text using term extraction and whitelist tagging while the data is live



Improved wizards for selecting gold-standard data. OBI Server as a self-service data source



ORACLE

# Discovery Application Lifecycle

## Building applications in days, not months !

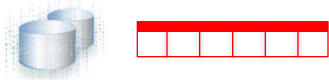
Diverse and changing information integrated and enriched via ETL

Automatically unified in Oracle Endeca Server – no predefined model required

Drag-and-drop application composition in Studio

Interactive search, navigation and visualization for exploration and analysis

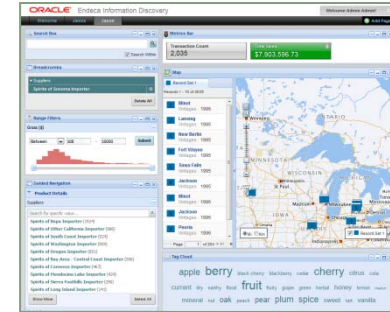
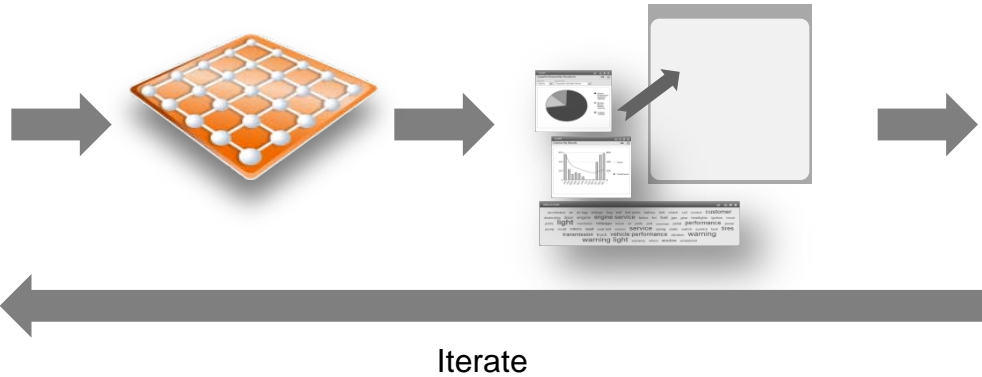
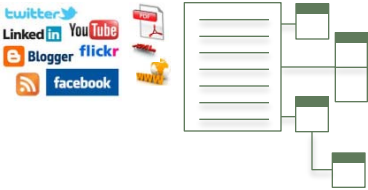
Structured



Semi-Structured



Unstructured



## Manufacturing



### Design

Discover the best parts and suppliers to use for your design and evaluate them for compliance, form, specification or any other attribute. Improve part re-use and approved supplier selection.

### Logistics

Gain insight into performance of aircraft, fleets, airports and airlines with an integrated view across geographic, weather, service and performance information. Analytics support early detection of patterns and opportunities for optimization.

### Automotive Claim Analysis

Explore and analyze any dimension and aspect of automotive warranty claim and recall data from the US National Highway Traffic Safety Administration (NHTSA). Text analytics and term extraction provide unique insight into claim history and trends.

### Sourcing

Bridge the gap between engineering and sourcing organizations to enable better and informed decisions on new designs and changes in engineering and effective spend and risk reduction across the supply chain.

### Warranty & Quality

Trace warranty claims and support issues back to the source supplier, plant and components to quickly address issues that can affect customer satisfaction and product quality.

## IT Support



### IT Asset Management (New)

Understanding the set of business assets that join financial, contractual and inventory functions to effectively rationalize and reduce IT expenditure, including all elements of software and hardware resources.

## CPG and Retail



### Marketing Insight

Map market reviews of your own and competitive products to help optimize campaigns and promotions. Incorporate the voice of the customer from product reviews and social media.

### Retail Analytics

How are your sales channels performing? Optimize your online and store channel marketing and sales promotion efforts by correlating point of sale data, customer demographics, reviews and product information.

### Sales

Analyze your customer and sales data to better understand performance of channels, brands and markets and evaluate the performance of incentives, promotions and sales.

### Social Media (New)

What are customers saying about your products or services? Gain insight into any dimension of social media feeds including Twitter, Facebook, newsfeeds and forum sites.

## Healthcare & Life Sciences



### Clinical Trials

Information on clinical trials provide a rich source of structured and unstructured data to mine for important trends and insight. Healthcare and Life Sciences professionals can explore across any dimension.

### Patient Care

Understanding details of patient care is a critical factor in optimizing and improving healthcare. Explore spending, admittance patterns, demographics and treatments to gain insight and make better decisions.

## Financial Services



### Financial Forensics

Gain insight into financial ledger information through unconstrained exploration. Support auditing and fraud detection through fully integrated search across structured and unstructured record information.

### Investment Research

Provide better customer support and optimized portfolio management through an integrated view for comprehensive and insightful investment research.

### Risk Management

Risk is measured by a wide range of factors and can vary greatly based on demographics. Explore and understand the risk of your loan portfolio and customer base across any customer, account or geographic dimension.

## Public Sector



### City of Chicago (New)

Explore and analyze any dimension and aspect of crime rates as well contract/payment data from the City of Chicago. Geospatial and search term analysis provide unique insight into crime history and trends.

### NY Financial Crime (New)

Prevent specific types of property fraud by analyzing any dimension and aspect of property related documents to detect unusual property activity and to find strong indicators of real estate fraud.

## Business Services



### Staffing

Make optimal decisions for global project staffing based on skills, location and availability. Improve utilization, client satisfaction and reduce outsourcing costs.

# Chicago Police Department Uses Discovery For Criminal Investigations

## Oracle Endeca Information Discovery



Chicago is the most populous city in the U.S. state of Illinois, and the third most populous city in the United States, with approximately 2.8 million residents. The Chicago Police Department (CPD), is the second largest local law enforcement agency in the United States. It has about 12,244 sworn officers and over 1,925 other employees.

### Before

- Objective is to help detectives with criminal investigations
- Difficult to efficiently combine and use data from crime incidents, case notes, suspect information, 311 call transcripts, etc
- Difficult today to do unscripted discovery to find:
  - For these types of crimes show me who the suspects are
  - For this suspect show me who is similar to this
- Short term objective: ability to better track social media data for a one-time event, NATO Summit in May 2012



### After

- Built real-time analytics app in the space of a few days to track twitter activity during the NATO Summit, understand patterns and better assign/deploy resources around the city.
- Incident Discovery is part of larger Big Data project
- Plan is that Discovery will help improve the ability of detectives to perform criminal investigations

# Toyota Reduces Time To Insight By 80%

## Oracle Endeca Information Discovery



Toyota Motor Corporation is a Japanese multinational automaker. Employing 300,000 people worldwide and with \$220B in revenue it ranks #8 on the Global 500 list.

### Before

- Need to find root cause of unintended acceleration problems, fast
- Needed to combine 20 data sources, structured and unstructured, internal and external
- Could not train engineers to write queries to sift through data
- Traditional app estimated to take over a year to build, unacceptable costs



### After

- Reduced time to insight by 80%
- Saved 100,000s of hours of engineer wait time
- EID system developed in just 12 weeks
- Helped prove no electronic-based acceleration issue
- Combine data from 20 sources , both structured (vehicle warehouse data, QA data, National Highway Transportation Safety Authority (NHTSA), etc.) and unstructured (internal warranty claims, externally filed complaints, etc), internal and external
- Easily searched, browsed, and aggregated

# Whirlpool Reduces Product Complexity and Improves Quality

## Oracle Endeca Information Discovery



Whirlpool Corporation is the world's leading manufacturer and marketer of major home appliances, with annual sales of approximately \$18 billion in 2012, 68,000 employees, and 65 manufacturing and technology research centers around the world.

### Before

- Growth through global acquisitions introduces complexity and leads to higher direct material costs
- Limited visibility to vendor activity between regions meant sub optimized supply agreements
- -Multiple ERP systems meant multiple vendor records within and between regions
- Multiple organizations maintaining vendor and item master data differently meant inconsistencies in data maintenance and control



### After

- Combined a dozen diverse operational systems and enterprise apps without having to perfect the data.
- -EID unified 4 regional SAP BW, 3 SAP R/3's with item master, part catalog, engineering change notices, engineering projects and "should cost" data.
- -Enables engineers to find the best part based on all salient technical, commercial and quality attributes. It provides commodity teams the ability to identify material cost reduction opportunities at a global level
- Buy only the parts it needs across all product lines, worldwide
- Target a 12% reduction on a multi-billion dollar spend
- Improved component reuse
- Reduced warranty and cost of quality
- Improved product innovation
- Reduced time-to-market



# US Army Improves Unit Readiness Evaluations

## Oracle Endeca Information Discovery



The United States Army (USA) is the main branch of the United States Armed Forces responsible for land-based military operations. It is the largest and oldest established branch of the U.S. military, and is one of seven U.S. uniformed services. The Force Validation Committee reviews force-management levels related to mission readiness.

### Before

- 40 Full-time action officers to evaluate Army units preparing for mobilization or deployment.
- Inability to link structured and unstructured information about a unit and its capabilities.
- Lack of visibility into Army Force Generation cycle:
  - Multiple data sources and formats
  - Departmental stovepipes
  - Data classification
  - Data integrity

### After

- Single system to query and drill down on to more than 20 major databases across the Army.
- Complete personnel view for rank, specialty, duty status and training level. Also includes unit readiness, location and equipment readiness.
- “The addition of Oracle’s Endeca Information Discovery product to the program provides that Army with an advanced, yet intuitive, exploration and analysis of complex and varied data across several functional areas such as readiness, manning, equipping, and training.”
- Personnel and Unit Readiness: Exploration and analysis of complex and varied data across several functional areas such as readiness, manning, equipping, and training



# What is Data Discovery?

Simplified

Quickly explore all relevant data



- ✓ Relationships undefined or unknown
- ✓ No pre-defined model required
- ✓ Rapid, iterative change



- ✓ Advanced search
- ✓ Faceted navigation
- ✓ Analytics



- ✓ Structured
- ✓ Semi-structured
- ✓ Unstructured
- ✓ Even messy data is ok
- ✓ Beyond the data warehouse

ORACLE®