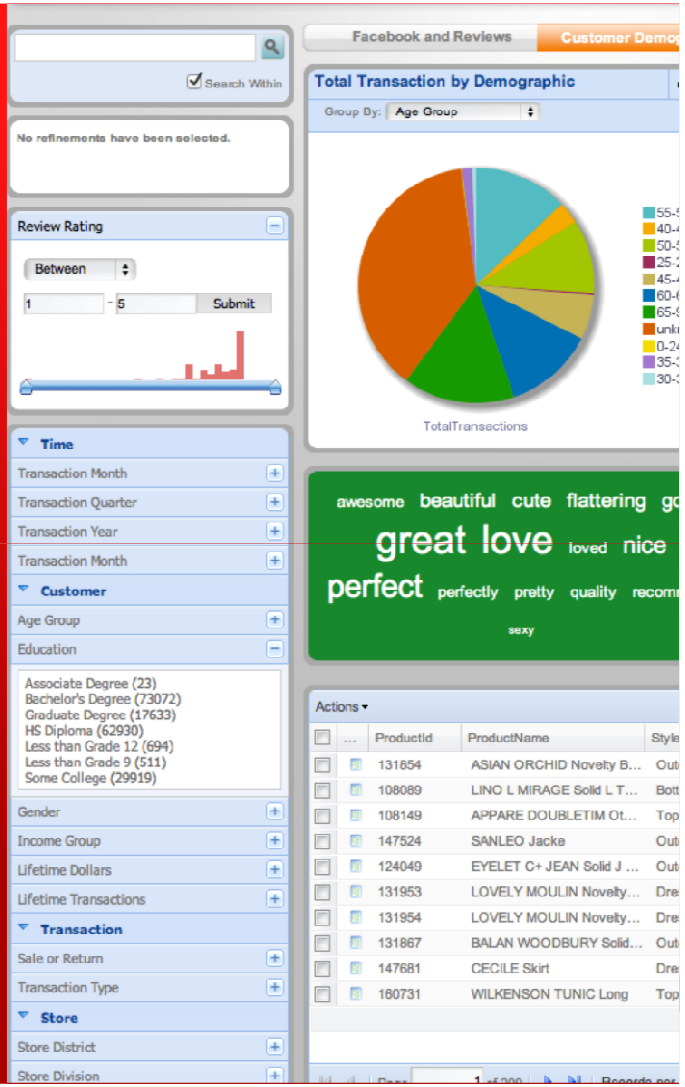
ORACLE

Oracle Endeca Information Discovery

The Missing Link

Ljiljana Perica, Oracle Business Solution Team Leader

Ljiljana.Perica@oracle.com



Data Discovery: Established Category for BI & Analytics

LICENSED FOR DISTRIBUTION

Gartner

Magic Quadrant for Business Intelligence and Analytics Platforms

5 February 2013 ID:G00239854

Analyst(s): Kurt Schlegel, Rita L. Sallam, Daniel Yuen, Joao Tapadinhas

Learn how Gartner can help you succeed

Become a Client now

VIEW SUMMARY

The dominant theme of the market in 2012 was that data discovery became a mainstream BI and analytic architecture. The market also saw increased activity in real time, content and predictive analytics.

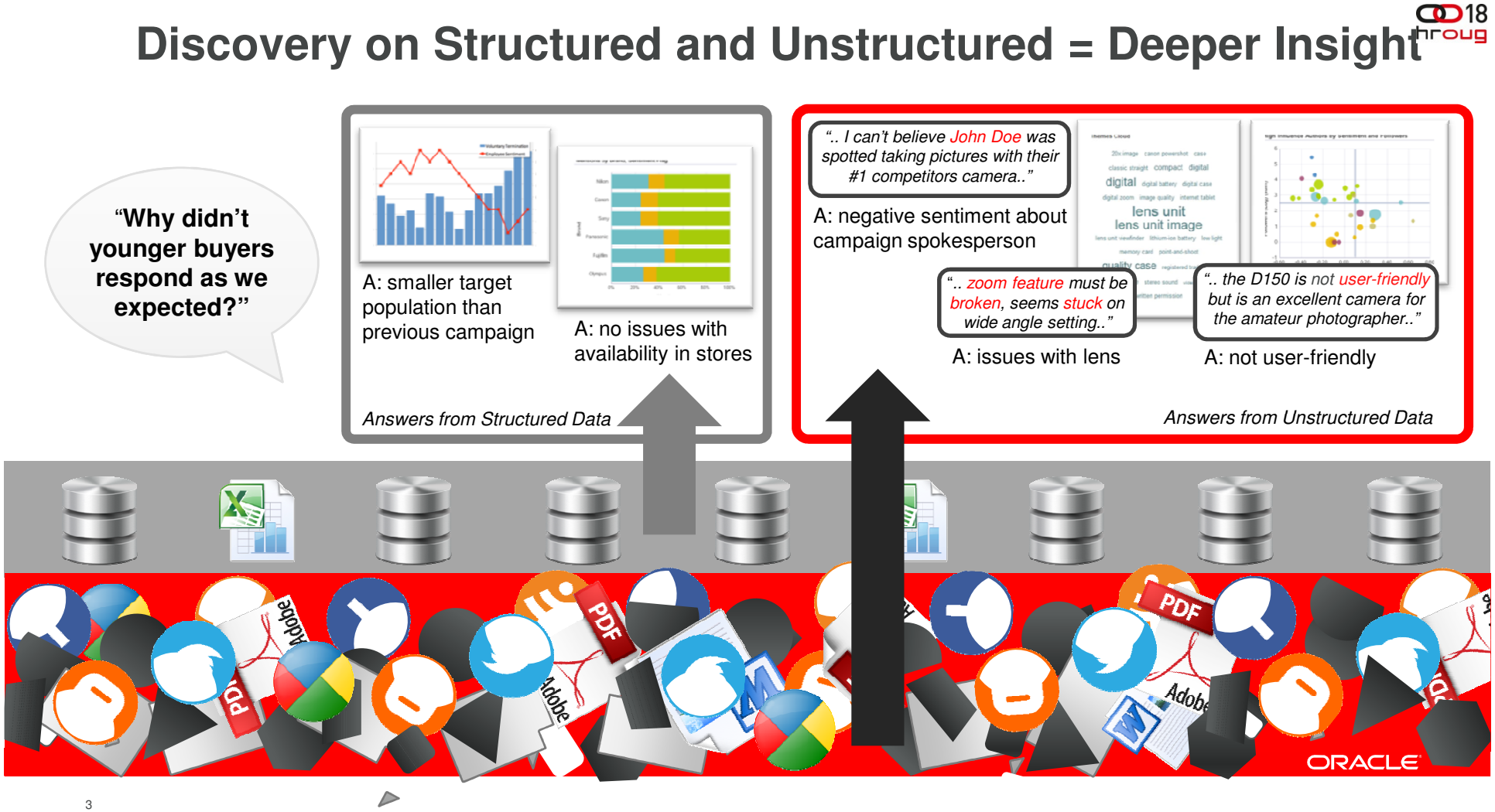
...UATION CRITERIA DEFINITIONS

... to Execute

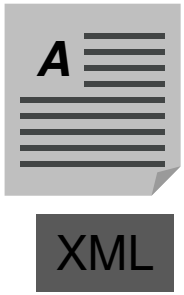
...uct/Service: Core goods and services offered by ...endor that compete in/serve the defined market. ...ncludes current product/service capabilities, ...y, feature sets, skills and so on, whether offered ...ely or through OEM agreements/partnerships as ...ed in the market definition and detailed in the

5 Feb 2013, Gartner MQ for BI and Analytics

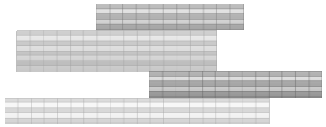
Discovery on Structured and Unstructured = Deeper Insight

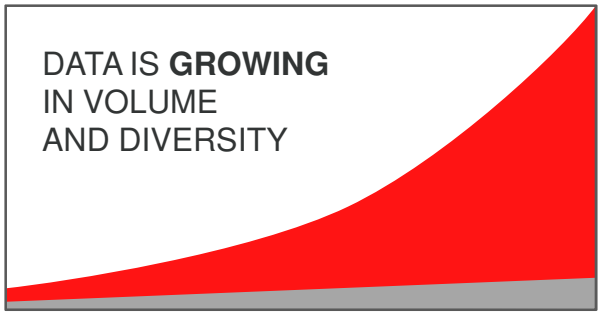


The Challenges of Unstructured Data



MOSTLY **TEXT**, AND
DIVERSE SCHEMAS

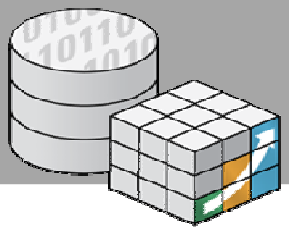




DATA CAN BE **DIRTY** OR
OF **UNCERTAIN VALUE**

— **20%** STRUCTURED —— **80%** UNSTRUCTURED

Business Intelligence and Data Warehouses



Text in Enterprise Applications

Enterprise Content Systems, File Systems, Email

Websites

Social Media

Big Data

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

Extend Business Analytics with Unstructured Data

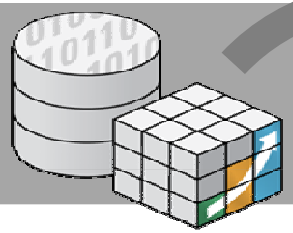
Introducing Oracle Endeca Information Discovery

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

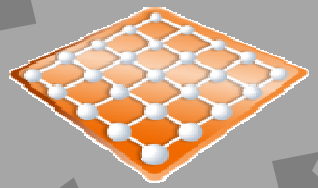
Best platform for integrated ROLAP and MOLAP

BI Server + Essbase
Common Enterprise
Information Model



Oracle Endeca Information Discovery

Best platform for Unstructured Analytics



Endeca Server
Hybrid Search/Analytical Database
Flexible Data Model

Structured Data

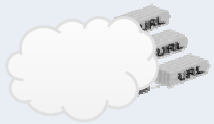
OLTP & ODS Systems

Enterprise Applications (Oracle, SAP, Others)

Data Warehouse & Data Marts

Unstructured Data

Websites



Content Systems, Files, Email

Social Media

Big Data

Brief Endeca History..

- Based in Cambridge, MA
- Founded in 1999
- 33% of the Fortune 100

ENDECA LATITUDE®

Data Discovery

ENDECA INFRONT™

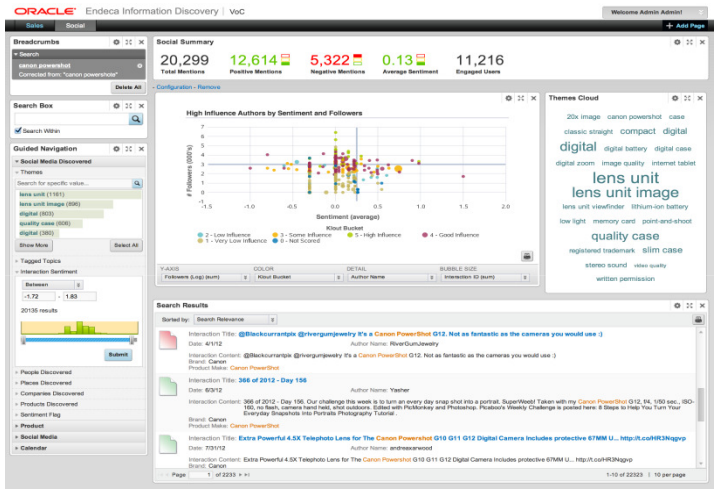
Customer Experience Management

Brief Endeca History: 2 Product Lines, 1 Core Technology

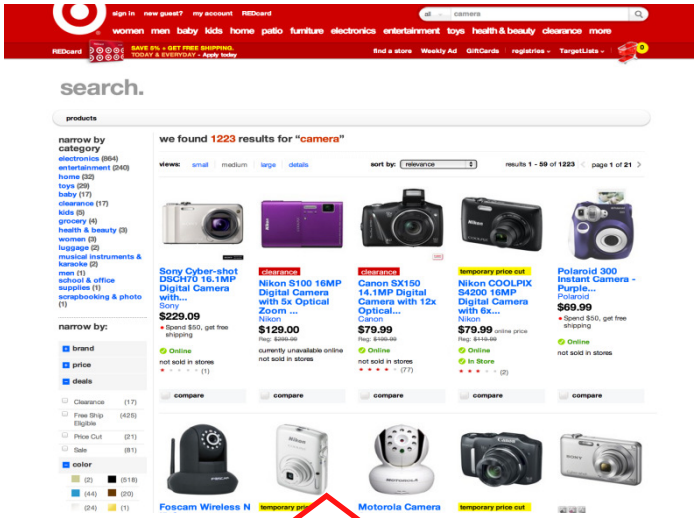
ENDECA LATITUDE → Oracle Endeca Information Discovery

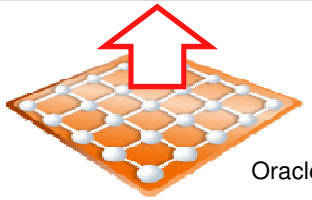
Oracle Commerce ← ENDECA INFRONT™

DATA DISCOVERY / BUSINESS ANALYTICS



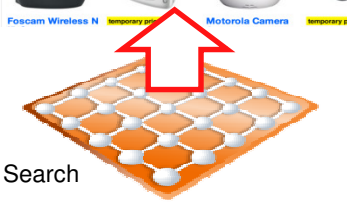
E-COMMERCE / CUSTOMER EXPERIENCE MANAGEMENT





MDEX Engine

Oracle Endeca Server ← → Oracle Endeca Guided Search



Analyst Proof Points

"**With this acquisition Oracle leapfrogged all other leading BI vendors in its capability to integrate unharmonized data sources and perform search based BI.**"

–Boris Evelson, Vice President, Principal Analyst

FORRESTER®

"**BI is arguably the greatest area of synergy and differentiation for Oracle from the acquisition of Endeca. It differentiates OBIEE from the products of traditional BI players, as well as from those of the emerging data discovery vendors.**"

– Rita L. Sallam, Mick MacComascaigh, Chris Fletcher

Gartner

"The combined products from Oracle's existing portfolio of information management, access, and analysis software (and systems) and Endeca features will give Oracle the scalability, the content+data analytical capabilities, the access and reporting tools, the repository, and the easy interactivity that are required to play in the big data arena."

– Susan Feldman, Research Vice President, Search and Discovery Technologies

IDC Analyze the Future

Oracle Endeca Information Discovery

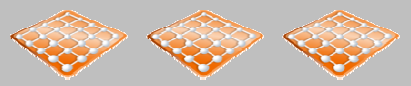
Platform Technology Overview

Studio

Intuitive Exploration and Analysis
Create and Share Apps

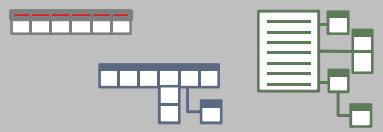
Endeca Server

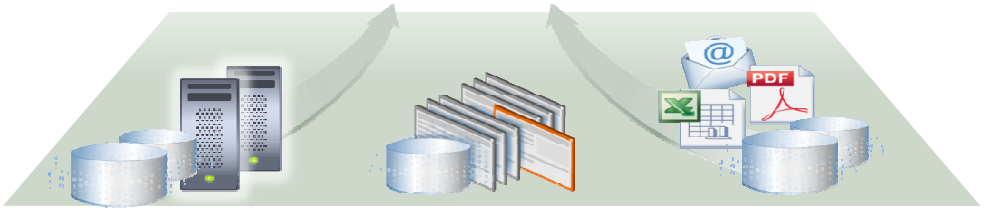
Hybrid Search/Analytical Database
In-Memory Architecture



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 enine & column-oriented in-memory database

Integration Suite – ETL

- Integrates Structured and Unstructured
- Text Enrichment and Sentiment Analysis

Discovery Application Lifecycle

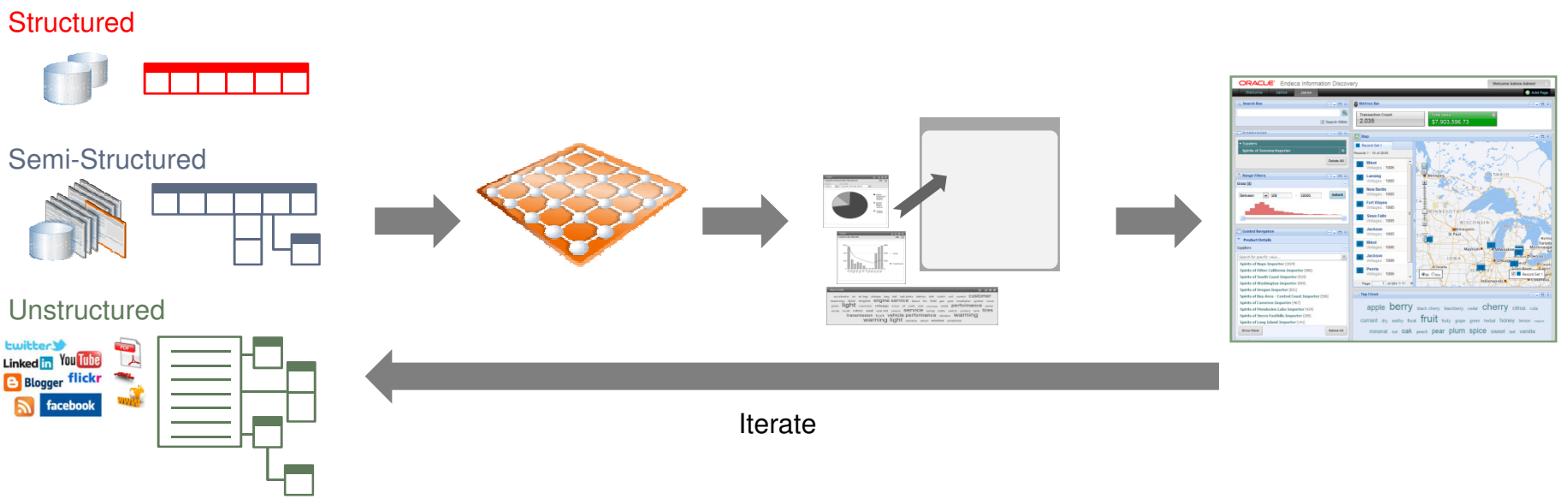
Building applications in days, not months !

Diverse and changing information integrated and enriched via ETL

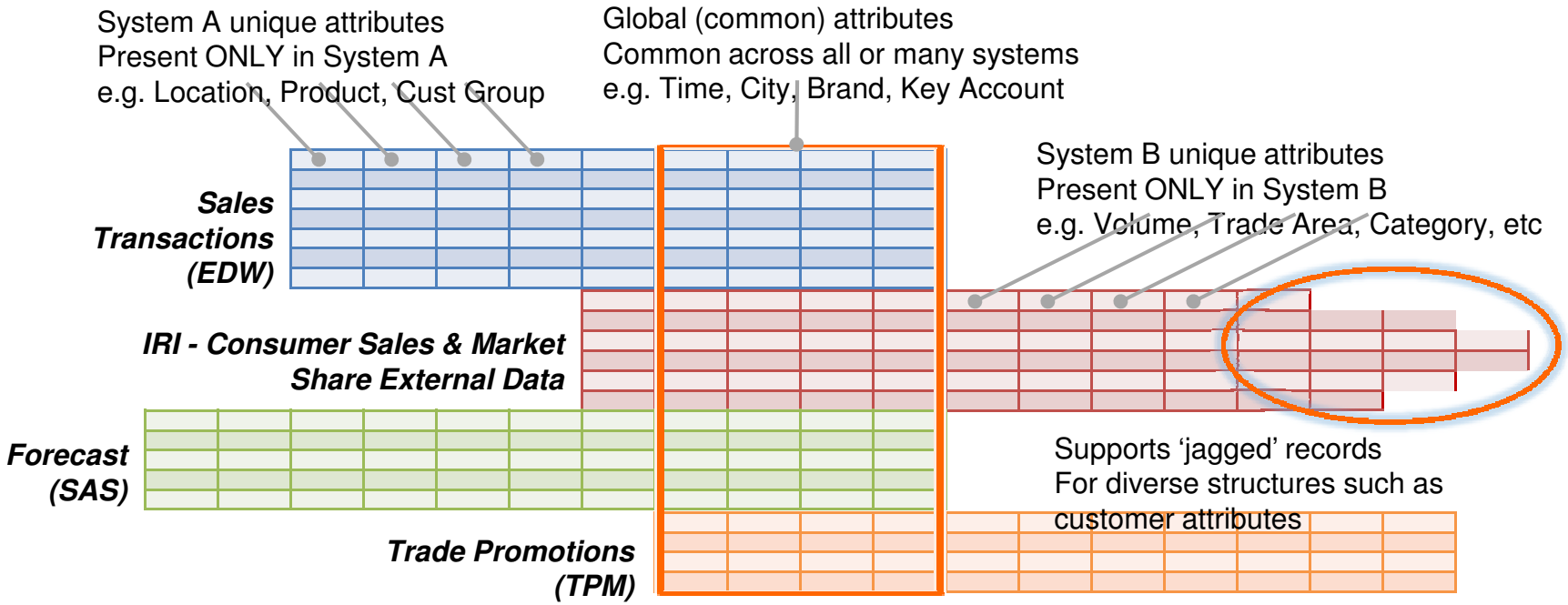
Automatically unified in Oracle Endeca Server – no predefined model required

Drag-and-drop application composition in Studio

Interactive search, navigation and visualization for exploration and analysis



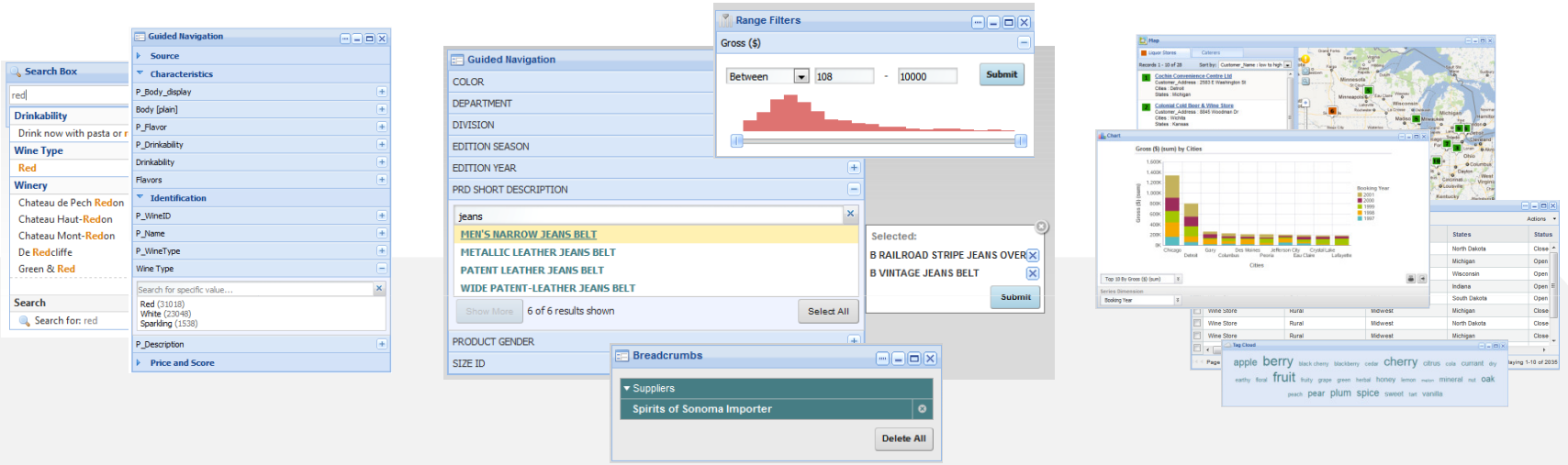
Endeca Server: Endeca brings any data together in a new way



- 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

Endeca Studio: Endeca's Unique User Experience

Interactive Data Exploration and Analysis



Deep Search

- Search across *all* data
- Dynamic typeahead
- Automatic spell correction
- Unlocks unstructured data

+

Contextual Navigation

- **Data-Driven.** Freely browse data without predefined paths or writing queries
- **Interactive.** Shows only valid next steps
- **Easy to Use.** Familiar online experience

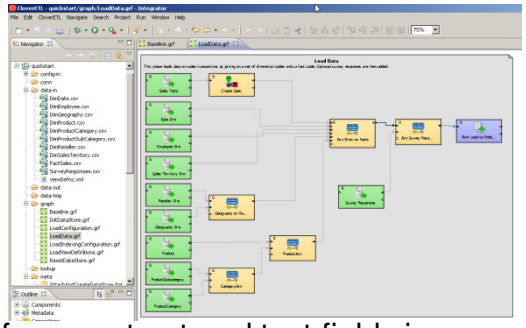
+

Visual Analysis

- Charts, crosstabs, key metrics
- Geospatial visualization
- Tag clouds

Endeca Integration Suite

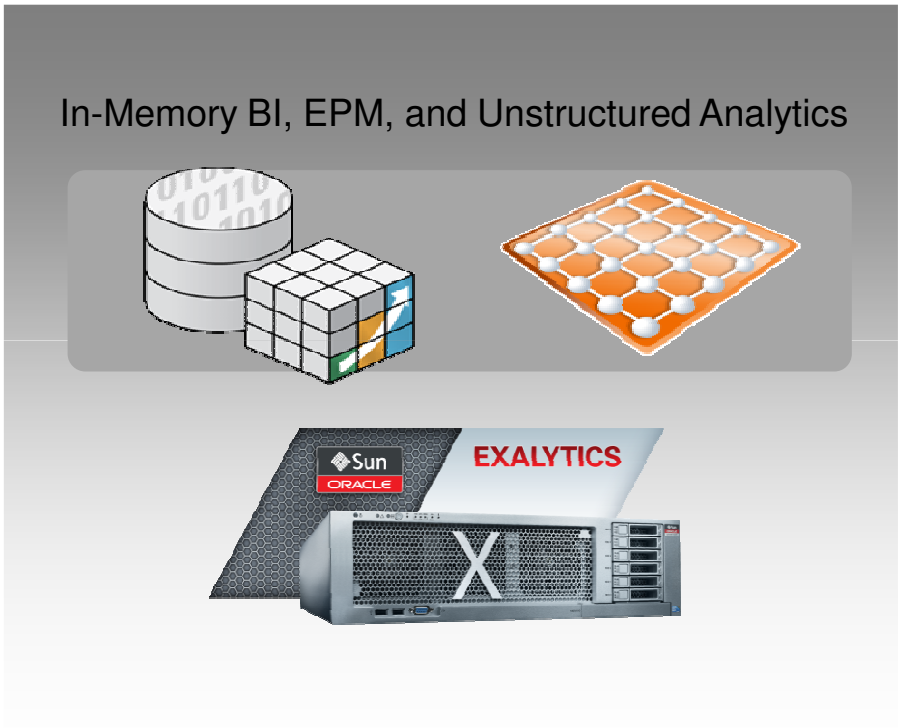
Loading diverse information into Endeca Server



- **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

Oracle Exalytics with Endeca Information Discovery

Fastest, most powerful hardware platform for Unstructured Analytics



- Endeca Software Breakthroughs
  - Multi-Core Parallel Query Evaluation
  - In-Memory Columnar Storage and Dynamic Cache
  - Adaptive In-Memory Data Mapping
  - Embedded Search Indices
  - Free Form Search, Exploration, and Analysis
- Optimized Hardware
  - Processor: 4 Intel® Xeon® E7-4870, 40 cores total
  - Memory: 2 Terabyte DRAM
  - Hard Disk: 3.6 Terabytes

ORACLE Endeca Information Discovery

Welcome! Log In. | Demo Home

Demo Home | How to... | Bookmarks | Sandbox | About...

Manufacturing

Automotive Claims

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

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

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

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

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.

Business Services

Staffing

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

Financial Services

Investment Research

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

Public Sector

City of Chicago

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

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.

I94 Visitors

Research specific types of applications by analyzing any dimension and aspect of I94 related documents to detect unusual activity and to find strong indicators of immigration violations.

© 2012 Oracle Corporation, All rights reserved

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, etc) and unstructured (internal warranty claims, externally filed complaints, etc), internal and external
- Easily searched, browsed, and aggregated

IBM Global Services Reduces Staffing Costs

Oracle Endeca Information Discovery

IBM Global Services is the world's largest business and technology services provider. It employs over 190,000 people across more than 160 countries.

Before

- Need to optimally staff 100,000 GBS and GTS consultants based on availability, bill rate, client and industry experience, skills, location, performance ratings and other unanticipated factors.
- Unable to easily combine data from PSFT HR, Siebel CRM, SAP scheduling and billing, Talent Management and Filenet Resume Store (full text)
- Unable to easily allow search on location, bill rate, level, industry experience, software experience, performance rating, languages, etc

After

- Reduced subcontractor utilization 5% to 7%, saved $500M year 1. Utilization rate (percent of time billed) has also risen.
- Applying principles of supply chain management to people, IBM created a database of employee profiles that managers use to match the right consultant with the right job
- Deployed to 6000 PMs, Managers globally
- 90 days from idea to production

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

Kraft Uncovers Root Cause Of Sales Spike in Unstructured Social Media Content

Oracle Endeca Information Discovery

With 25,000 employees in the United States and Canada, Kraft Foods is North America's 4th largest consumer packaged food and beverage company, with revenues of approximately $19 billion

Before

- Within consumer marketing group, wanted to better understand the consumer so can more effectively market to them and sell more products
- Found traditional data sources were often unable/insufficient to explain variations in sales and performance in product and brand (looking at pricing and promotion).
- Difficult to combine non-traditional data sources, like Social Media, customer service call reports, customer satisfaction surveys, to help investigation

After

- Discovered a signal – a root cause of an otherwise unexplainable spike in sales - out of noisy data
- Combined structured data (internal data for all orders, AC Nielson data for promos, for ad spend) with unstructured data from call center transcripts in CRM and social media data in Radian6
- Successfully found root cause of a spike in Macaroni and Cheese sales, related to YouTube viral video of new voice-over artist for the product.

Oracle Endeca Discovery DEMO

Questions

ORACLE®