Oracle Endeca Information Discovery
Overview and Frequently Asked Questions

Overview

Most organizations have implemented business intelligence to enable decision makers to monitor, understand and improve business performance.

Yet over 80% of enterprise information exists as unstructured data beyond the reach of traditional BI environments - in websites, social media, content management systems, emails, documents, sensor data, external databases, and more - and this diverse and changing data is growing exponentially. Organizations that cannot understand the insights buried within this data are at a critical competitive disadvantage.

Oracle Endeca Information Discovery solutions unite the worlds of structured and unstructured data to provide business users with complete visibility into their business processes, creating new insights and enabling better business decisions. Only Oracle provides a complete, integrated solution for superior structured and unstructured data analysis, all delivered on engineered systems optimized for performance and low total cost of ownership.

Oracle Endeca Information Discovery helps solve problems associated with analyzing unstructured data, in finding better ways to listen to customers, improving the relationship between companies and their employees, getting to the real root causes behind issues, or creating new products and services through a better understanding of the market. Oracle Endeca Information Discovery creates transformative opportunities for the business, and enables IT to cost effectively support them.

Key Capabilities

- **Intelligent Search** Use intuitive, yet advanced, search for rapid identification of relevant business concepts and instant results across all your data. Search helps users find what they need, even if they don't know how to ask for it.
- **Interactive Visualizations** Visualize data in context to quickly investigate root causes, identify exceptions, and surface new concepts.
- **Contextual Navigation** Originally built for and proven on leading consumer web sites, interactive faceted navigation helps users quickly make sense of unfamiliar data sets.
- **Universal Data Support** Endeca’s unique hybrid search/analytical database and comprehensive library of connectors for structured and unstructured sources support integration of any data, regardless of schema or type.
- **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.
- **Drag-and-Drop Application Composition** Creating compelling user interfaces has never been easier with an interactive, drag-and-drop authoring environment for quickly assembling enterprise-class discovery applications.
- **Dynamic Metadata Layer** An optional, adaptive semantic layer that allows incorporation of business definitions from existing BI tools without getting in the way.
- **In-Memory Architecture** Highly scalable, column-oriented, multi-core parallelized, and natively in-memory Endeca is built to optimally leverage engineered systems like Oracle Exalytics In-Memory Machine to deliver speed-of-thought performance.

Frequently Asked Questions

Overview

What is Oracle Endeca Information Discovery?

Oracle Endeca Information Discovery is an enterprise data discovery platform for rapid, intuitive exploration and analysis of data from any combination of structured and unstructured sources.

Why Oracle Endeca Information Discovery?

Oracle Endeca Information Discovery enables organizations to extend their existing business analytics investments to unstructured sources – including social media, websites, content systems, files, email, database text, and Big Data - providing unprecedented visibility into data and business processes, saving time and cost, and leading to better business decisions.

With Oracle Endeca Information Discovery, customers can:

- Easily combine structured and unstructured data and metadata to understand key metrics in their relevant context and evaluate new business situations.
- Ask unanticipated questions of any data through intuitive, flexible, and highly interactive online discovery applications.
- Mash up and recombine data and visualizations to compose discovery applications.
- Leverage all of the rich metadata of their data models and the semantic layer as a basis upon which to build discovery applications.
- Unburden IT from the constant chasing of new requirements and nontraditional data sources for incorporation into the data warehouse, enabling them to deliver fast access to relevant data and self-service to business users while maintaining security, governance, and quality.

What are examples of how customers use Oracle Endeca Information Discovery?

Many Oracle customers have successfully complemented their existing business analytics investments with Oracle Endeca Information Discovery. Here are three examples:

Automotive Manufacturing

Several years ago a large automotive manufacturer issued a massive vehicle recall related to reports of unintended acceleration leading to several deaths. While the CEO was called before Congress to explain the situation, they faced fundamental questions: “Is this a real quality problem, or something else? How exposed are we if it is a quality issue? What are our customers saying about it and how is it affecting our sales?”

The company is a very happy Oracle Business Intelligence customer, but there were no reports to answer these questions. Using Oracle Endeca Information Discovery, they were able to combine a variety of data from their warehouse and beyond – vehicle data, quality reports, internal warranty claims, sales transactions, service records, supply chain data, and more. When new questions required data from outside the company they were able to readily incorporate claims from the National Highway Transportation Safety Authority and competitor sales data from JD Power. Only by combining all of this data – replete with misspellings and bad grammar – did they have the right infrastructure in place to enable line of business workers to understand what was happening.

The quality engineers, the marketing organization, and the team managing the supplier relationships have the expertise to ask questions about vehicles, suppliers, manufacturing processes and facilities, but they don't have the expertise to write advanced queries or build reports. Oracle Endeca Information Discovery enabled these business users to easily explore, analyze, and understand this diverse data.

After a thorough investigation, the company was vindicated. The Transportation Secretary concluded there was no electronic-based cause for unintended high-speed acceleration in their cars. Proving a negative – that the cars didn’t have an electronic problem – was tough. And the Oracle Endeca Information Discovery played a prominent role in exonerating the company.

The company estimated that it would have taken over a year to solve this problem with their traditional BI tools. EID reduced time to market by 80%. The company also estimates that the engineers' ability to ask and answer their own questions as they unfolded through the investigation saved hundreds of thousands of hours engineers would have spent waiting for reports to answer their new questions.

Consumer Beverages

A major consumer beverages company needed to understand variances between demand forecasts and actuals. While this is typically a problem well served by business intelligence tools, their demand planners still had additional questions based on the need to understand why inaccuracy existed in the demand plan. They wondered: "Could variations be due to unanticipated trade promotions with customers? Does pricing impact the accuracy of the demand plan? What about unanticipated shipments of products between distribution centers?"

They built a discovery application for the demand planners that combined the forecasts out of SAS with the actuals from the distribution transaction system, and then connected a separate marketing database with the other two sources. When they saw that some of the variances were still unexplained the planners had more questions: "Do promotions offered by our sales team lead to unanticipated bulk buying?", To address this they loaded Trade Promotion data from outside the data warehouse. Then the planners asked: "Did our customers affect demand by changing their prices? Did competitor pricing impact demand?" They then combined sales and pricing data acquired from 3rd party sources. All of this happened in 8 weeks.

Finally, planners discovered something they didn't expect. When they asked the question, "How do out-of-lane shipments between distribution centers impact forecast accuracy?", they actually found that unauthorized overrides to the demand plan being performed by individuals in the field was actually help to improve accuracy of the forecast. This was due to tribal knowledge of business conditions, impossible to predict in the planning process. These tribal business practices have now been captured and replicated across the business leading to accuracy improvements of between 2-5%.

Commercial Food Production

The world population hit 7 billion last year. A large processed food producer knows that corn yields have to go from about 150 bushels an acre to about 200 to feed a growing world. One division sells and distributes new strains of seed to increase farmers' crop yields. But first, they have to get the farmers to use the seed. Farmers are no-nonsense people. And they often can't weather even a single season of poor yields. So the food producer has to make the case with data.

Fortunately, there is lots of data available. The trick is combining it cost-effectively and making it usable and useful. Oracle helped this food producer combine data from many sources including a transactional warehouse that indicated which farmer had bought what, a marketing database that indicated which farmer had been pitched what seed, and a separate transactional warehouse with data from "answer plots" that the company plants all around the US at different latitudes in different soils with different seeds to demonstrate the actual yields. Finally, all this was combined with government data on how many acres are planted with which crops. Data from these multiple sources, some of which are outside the company's control and can change at any time, were combined to derive insights.

This application is now used by hundreds, and soon thousands, of salespeople, many of them former farmers. The company expects higher profit margins as a result. But already they estimate they saved 1.5 years and $4M by solving this problem with Oracle Endeca Information Discovery.

How does Oracle Endeca Information Discovery differ from Oracle Endeca Commerce?

In December 2011, Oracle acquired Endeca Technologies, Inc., a leading provider of unstructured data management, web commerce and business intelligence solutions.

Endeca InFront, now part of Oracle Commerce, is a leading customer experience management platform that enables businesses to deliver highly targeted and relevant customer experiences online with advanced merchandising and content targeting tools for web commerce.

Endeca Latitude, now called Oracle Endeca Information Discovery, is a technology platform that enables businesses to

enhance existing analytic investments by bringing information from many unstructured and structured information sources together.

Unstructured

What is unstructured 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 is text-heavy and generally in free-form text. A few examples of this type of data are email, documents, presentations, web content and social media. Due to the explosion and proliferation of the internet and social media, unstructured data is growing exponentially – and companies are looking for better ways to manage this data.

In contrast, structured data can be organized and is easily identifiable. Users can search by data type within the actual content. For example, SQL (Structured Query Language) allows user to select specific pieces of information based on columns and rows in a field. Structured data has traditionally been captured and stored by enterprises. For example, information generated by enterprise resource planning and supply chain applications is structured data.

How does Oracle Endeca Information Discovery combine structured and unstructured data?

Oracle Endeca Information Discovery makes it easy to extend existing business analytics with unstructured data from beyond the warehouse through Oracle Endeca Server's flexible data model and metadata layer.

One of the main differentiators of Endeca Server is its flexible data model. as described above, Endeca Server's key-value store enables the rapid integration of any type of data, whether structured or unstructured, from inside or beyond the confines of the enterprise data warehouse.

Oracle Endeca Server also includes a unique flexible metadata management capability – the View Layer – that enables direct incorporation of semantic definitions from customers' existing structured BI environments for rapid time to analysis. This enables business users to understand familiar business metrics in the context of additional information from unstructured sources.

Why is Oracle Endeca Information Discovery perfect for social media and other web content?

Oracle Endeca Information Discovery is perfectly suited to acquire, store and analyze diverse and unstructured data from modern online sources, such as social media sites, blogs and forums, news sites, consumer reviews, and public data sets. These sources present three serious challenges for traditional analytics tools: the data is often only accessible via a web service, API, or content crawl instead of traditional SQL; most of the valuable content is human language in text fields or spread across an array of metadata fields; and each source may have diverse and changing schemas. Because this data is inherently unstructured and unpredictable, it's difficult for users to make sense of it. Endeca is designed from the ground up address these challenges by: providing connectivity and text enrichment for a wide variety of structured and unstructured sources; managing content in a key-value store that supports any data type without requiring a universal schema; and providing an intuitive, interactive user interface for searching, exploring, and analyzing this diverse content to quickly surface valuable new insights.

What is text enrichment?

Text Enrichment is the 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 provides this capability, along with basic term and regular expression whitelist tagging.

What is sentiment analysis?

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.

What technologies does Oracle Endeca Information Discovery use for text enrichment and sentiment analysis?

Oracle Endeca Information Discovery employs the Lexalytics Salience engine to perform text enrichment and sentiment analysis. The integration is implemented as a data manipulator component in Integrator, the out-of-the-box ETL tool.

What content management system connectors are available?

CMS connectors are available for Documentum, Documentum eRoom, Filenet Doc & Image Svs, Filenet P8, Interwoven TeamSite, JSR-170, LiveLink, Lotus Notes/Domino, and Microsoft SharePoint.

What additional skillsets are needed to use Oracle Endeca Information Discovery?

The skills required to support an Oracle Endeca Information Discovery deployment are similar to those supporting BI deployments. Typical projects include an ETL developer, data/solution architect, and a team of business analysts, in addition to project managers, stakeholders and end users. System administrators' time is required to set up the software environment and ensure it is ready for production use. A difference from typical BI implementations is that iterative Oracle Endeca Information Discovery deployments can be up in days and go live in weeks to months.

Architecture

What are the key technologies behind Oracle Endeca Information Discovery?

Oracle Endeca Information Discovery consists of three major components: Endeca Server, Studio, and Integrator. Endeca Server is a unique hybrid search/analytical database designed for enabling interactive exploration and analysis of diverse and unstructured data. Studio is the web application that serves as the user interface for business analysts to quickly assemble interactive component-based applications and for end users to explore and analyze data. Integrator is the provided ETL tool that loads structured and unstructured source data into Endeca Server; because Endeca Server is an open platform, customers may also use other ETL tools that provide web services connectivity.

What type of analysis is Oracle Endeca Information Discovery designed for?

The core use case for Oracle Endeca Information Discovery is in supporting a users' need to better understand the data. In expanding analysis inputs beyond structured content to unstructured sources, users can better understand the color and qualitative insights which support quantitative results.

Oracle Endeca Information Discovery enables this type of analysis through Oracle Endeca Server, a hybrid search/analytical database that is specifically optimized for discovery, not reporting or online transaction processing. Highly scalable, column-oriented, and in-memory - without being memory bound - Oracle Endeca Server supports navigation, search, and analysis of any kind of information including structured, semi-structured, and unstructured content.

What is a hybrid search/analytical database?

Combining an index-based search engine and column-oriented, in-memory analytical database in a single integrated architecture, Oracle Endeca Server was designed from the ground up to meet the needs of an exploratory data discovery user experience across any combination of structured and unstructured data.

How is data represented in Oracle Endeca Server?

One of the main differentiators of Endeca Server is its flexible, self-describing data model. There is no need for IT to define a unified schema before loading and analyzing data. Unlike a relational model, it is a key-value store – a collection of records that each contains its own, potentially unique, collection of attributes – that enables the rapid integration of any type of data, whether structured or unstructured, from inside or beyond the confines of the enterprise data warehouse, without the efforts associated with traditional relational data modeling. The data is not segmented into tables nor is there a

universal schema to which all records must conform. This architecture easily accommodates:

- Idiosyncratic structure. Records are self-describing, each representing its own possibly unique schema.
- Multi-valued fields. Fields can have multiple values, creating new possibilities for data representation.
- Unstructured fields. Native support for text fields of any length.

Does Oracle Endeca Information Discovery require data modeling?

Yes, but it is a substantially different process than traditional relational or dimensional tools. Oracle Endeca Server does not have a pre-defined or rigid data model, instead, it starts as a "blank slate" without data or schema. Data may be added at any time via web services, at which point Oracle Endeca Server examines each ingested record and dynamically creates new attributes in the schema if those attributes have not yet been seen on other records. Each ingested row becomes an “Endeca record”, with each record being uniquely defined by a customizable “record specifier”, akin to a primary key in a database. The model (e.g. schema, data grain) is therefore defined by what you choose to ingest. You also have the option to manually setup the schema and define record specifiers if you would prefer to explicitly control the configuration.

What metadata management capabilities does Oracle Endeca Information Discovery provide?

Oracle Endeca Server also includes a unique flexible metadata management capability – the View Layer – that enables direct incorporation of semantic definitions from customers’ existing structured BI environments for rapid time to analysis. This enables business users to understand familiar business metrics in the context of additional information from unstructured sources.

Unlike traditional semantic layers, it supports incremental definition of metrics, dimensions, and other metadata after data has been loaded into an application. This enables customers to quickly explore and analyze new disparate data sets without spending time up front building a semantic model; users can iteratively grow metadata as understanding of the data matures.

What types of visualizations does Oracle Endeca Information Discovery provide?

Oracle Endeca Information Discovery includes a wide variety of drag-and-drop visualization and filtering components, including: Alerts, Bookmarks, Breadcrumbs, Chart, Compare, Crosstab, Data Explorer, Guided Navigation, Map, Metrics Bar, Range Filters, Record Details, Results Table, Results List, Search Box, and Tag Cloud.

What search capabilities does Oracle Endeca Information Discovery provide?

Oracle Endeca Information Discovery provides two different, and complementary, types of search:

- Value – or dimension – search provides type-ahead auto-completion that enables users to find attribute values and help disambiguate queries by identifying which attributes contain the search terms. For example, a search for “Houston” may return “City > Houston, TX” and “Company > Houston Drilling Co.”, giving the user the option to choose which she meant. Selecting a result applies that attribute value as a filter to the results.
- Record search provides advanced full text search across any single or combination of fields, leveraging a variety of configurable term matching and relevancy ranking algorithms to retrieve results for display and analysis.

Additional out-of-the-box features like automatic spell correction, “did you mean?” suggestions, stemming, and synonym matching ensure that users can find what they need, even if they don't know how to ask for it.

What languages does Oracle Endeca Information Discovery support?

Oracle Endeca Information Discovery (EID) has been optimized to support ingest, display, full search, navigation and text enrichment for the English language. In addition, EID provides language packs for French, German, Spanish, Italian,

Dutch and Portuguese enabling search capabilities. Currently, all product interfaces are provided in English only.

Does Oracle Endeca Information Discovery support enterprise security policies?

Yes. Oracle Endeca Information Discovery is built to seamlessly integrate with the existing security infrastructure and quickly extend data governance and security policies. Authentication is made easy for users (as well as IT) since it can integrate with a variety of standard enterprise authentication systems, including Active Directory, LDAP, and Single Sign On (SSO) and also support secure encryption of its network communications through SSL.

Equally important is the ability to restrict access to different levels of data. Access to information can be restricted by application, view (page/component) or data type (row level).

Oracle BI and Oracle BI Foundation Suite

Is Oracle Endeca Information Discovery part of Oracle BI Foundation Suite?

No. Oracle Endeca Information Discovery is a separately licensed product.

How does Oracle Endeca Information Discovery complement my existing BI investments?

Oracle Endeca Information Discovery extends and leverages these investments to incorporate unstructured data, adding important contextual information to business analysis. Users leverage all of the rich metadata of their data models and the semantic layer as a basis upon which to build discovery applications. An example usage scenario would be, combining warehouse sales data with Twitter, Facebook, warranty claim, and other customer feedback data for new types of market analysis.

We don't run Oracle BI. Can we still benefit from Oracle Endeca Information Discovery?

The diversity of supported data sources is a strength of Oracle Endeca Information Discovery. It supports all the types of data supported by Oracle Business Intelligence Foundation Suite, including DW solutions, OLAP, MOLAP, OLTP solutions, as well as unstructured sources. Oracle Endeca Information Discovery excels in providing "speed of thought" exploration and analysis across such diverse sources. Oracle Endeca Information Discovery can provide the benefits of data discovery with or without the foundation of Oracle BI.

Can we use Oracle Endeca Server as a data source for my existing BI tools?

Yes. Oracle Endeca Server exposes web services for querying its data. Note that these services are designed for many smaller interactive queries over filtered data, rather than large-scale bulk export of entire data sets.

How does Oracle Endeca Information Discovery leverage Oracle BI Applications?

Oracle Endeca Information Discovery enhances existing business intelligence applications by extending analysis to unstructured sources and supporting exploratory data analysis. Customers get to leverage all of the rich metadata of the BI applications as a basis upon which to build discovery applications.

Exalytics

Is Oracle Endeca Information Discovery compatible with Exalytics?

Yes, Oracle Endeca Information Discovery is certified on Exalytics. Exalytics is the ideal hardware for Oracle Endeca Information Discovery's CPU and memory-intensive workload.

What are the benefits of running Oracle Endeca Information Discovery on Exalytics?

Exalytics provides an ideal platform for Oracle Endeca Information Discovery, with 40 cores of processing power and 1TB of RAM. With Oracle Endeca Server's multithreaded parallel query evaluation, in-memory column storage, and dynamic in-memory cache, it can take full advantage of this state-of-the-art hardware.

Can I run Oracle Endeca Information Discovery and Oracle BI Foundation Suite on the same Exalytics machine?

Yes. Depending on the size of the implementation, customers may wish to either explicitly partition the Exalytics machine to control resource allocations, or purchase additional Exalytics machines in order to ensure each platform has sufficient resources for the workload.

Big Data

What is Big Data?

Big Data has several definitions that generally encompass at least three characteristics: Volume (the amount of data), Variety (the diversity of sources and structures) and Velocity (the speed of data change). Some definitions include a fourth characteristic: Value (a nod to the fact that most Big Data sources are of uncertain informational value and therefore companies are unsure what to do with it).

What benefits does Oracle Endeca Information Discovery provide in the area of Big Data?

Oracle Endeca Information Discovery's primary benefit with respect to Big Data is in solving the problem of Big Variety – relevant data sources for business analysis have never been more diverse, and an increasing share of this information is unstructured. Oracle Endeca Information Discovery combines and exposes the relationships between these data in new ways, guiding users to uncover new insights.

How does Oracle Endeca Information Discovery compare to Hadoop?

Oracle Endeca Information Discovery and Hadoop are highly complementary technologies. While Hadoop is designed for large scale distributed management and batch processing of huge data unstructured volumes, and Oracle Endeca Information Discovery is designed for interactive end user exploration and analysis of structured and unstructured data, they both share a similar flexible data model. This makes it straightforward to combine subsets of data from Hadoop with other diverse data sources in Oracle Endeca Information Discovery.

Oracle Corporation
Worldwide Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.
Worldwide Inquiries
Phone: +1.650.506.7000
+1.800.ORACLE1
Fax: +1.650.506.7200
oracle.com

Oracle is committed to developing practices and products that help protect the environment

Hardware and Software, Engineered to Work Together