ORACLE®
BUSINESS INTELLIGENCE

An Oracle White Paper
January 2014

From Specification to Exploration:
Data Discovery and the Transformation of
Business Analysis

ORACLE®

Contents

Executive Overview

How can we harness the potential of Big Data? With the proliferation of new sources of digital information, organizations are leaving a wealth of knowledge untapped. Unlocking the insights from new sources of information requires expanding the boundaries of traditional Business Intelligence systems. Data Discovery offers easy exploration across a big variety of data to provide business users with extensive new visibility into business performance. This new capability is beginning to transform the way that business analytic systems are provided and used. Instead of a lengthy process of specifying requirements for the system, Data Discovery allows rapid, intuitive exploration and analysis of information from any combination of sources. It enables deeper insight into the business and the opportunity for greater efficiency in the Business/IT relationship. It is bringing about the need for new roles, new leadership and new ways of thinking about data governance.

Introduction

Several years ago, a large automotive manufacturer issued a massive vehicle recall related to reports of unintended acceleration, leading to several deaths. While the firm knew it was being accused, it didn't know if there was actually a problem with the cars. No company has reports for problems it never expects to happen.

Using Oracle Endeca Information Discovery, the firm combined a variety of data from its warehouse and beyond—vehicle data, quality reports, internal warranty claims, sales transactions, service records, supply chain data, and more. It also included data from outside the company, like claims from the National Highway Transportation Safety Authority and competitor sales data from JD Power. This capability enabled the quality engineers, marketing organization and supplier alliance managers to easily explore and understand this diverse data without having to build advanced queries or reports.[1]

To remain relevant in the market and to stay ahead of competition, organizations are beginning to use new, diversified information regardless of source, to explore new paths, ask questions and validate ideas. Data Discovery enables companies to extend their existing business analytics infrastructure to unstructured data, resulting in improved business decisions based on a deeper understanding of the business. [2]

The Data Discovery process, and new tools like Oracle Endeca Information Discovery, are changing the nature of work, re-defining roles and re-aligning the traditional relationship between IT and Business. Organizations that embrace the new paradigm realize productivity increases on both sides; those that fail to adjust will become laggards in developing business insight and cross functional productivity. This paper has three parts: an introduction to Data Discovery, a review of the traditional software deployment process, and an exploration of the ways that Data Discovery is bringing about changes in the process, roles and impact of business analysis.

Data Discovery

Most organizations have implemented some form of Business Intelligence (BI) to enable decision-makers to monitor, understand, and improve business performance. Yet the vast majority of today's data growth is in systems beyond the reach of traditional BI environments—in websites, social media, content management systems, e-mails, documents, sensor data, external databases, and more—and this diverse and changing data is growing exponentially.

The type of data is varied, ranging from structured database tables, to semi-structured forms containing a mix of numbers and free-form text, to wholly unstructured documents. In fact it is the variety rather than the volume of data, and its uncertain value, that presents the biggest challenge. These factors, coupled with an internet-savvy business culture and the impact of consumer mobile applications on enterprise software have created an urgent need for business users to have the freedom to easily and quickly explore the relevant information to discover new insight. [3,4]

The process of quantifying data from all types of sources, in all types of formats is what co-authors Kenneth Cukier and Viktor Mayer-Schönberger call *datafication*.[5] Datafication allows information to be collected, tabulated and analyzed, so that the potential uses of the information are limited only by the ingenuity of the user.[6] "Data's true value is like an iceberg floating in the ocean. Only a tiny part of it is visible at first sight, while much of it is hidden beneath the surface. Innovative companies that understand this can extract that hidden value and reap potentially huge benefits." [7]

Enterprise-class data discovery systems enable rapid, intuitive exploration and analysis of data from any combination of structured and unstructured sources. They let organizations extend their existing business analytics investments to new combinations of a greater variety of sources – including social media, websites, content systems, email, and database text – providing a new level of visibility into data and business processes, saving time and cost, and leading to better business decisions. Some of the main benefits of this approach include: [8]

• Increased insight and visibility, enabling users to find the data they want to analyze, from which they can gain deeper insight into the business.

• Access to fresher information. Data and content can be delivered in near real time, helping people make decisions based on the most current information.

• Increased reuse of assets. Users are better able to find and reuse information assets, eliminating the costs of re-creating these assets.

• Consumer-style ease of use. Oracle Endeca Information Discovery, for example, lets BI professionals develop and deliver analytic applications that business professionals will actually want to use—leading to higher adoption rates, lower training costs, and faster time to value.

Data Discovery is finding uses in a rapidly growing range of industries and lines of business. For example, in Human Resources, monitoring employee satisfaction and sentiment or matching candidate skills with open positions. In Finance, combining customer sentiment with sales trends to predict financial performance. In the Public Sector, identifying potential disease epidemics and crime trends. In Sales and Marketing, spotting consumer product trends and hot themes before they begin to show up in sales results. In Manufacturing, understanding product or service satisfaction through a combination of warranty results, surveys and social media feeds.

Before proceeding to examine how Data Discovery changes the nature of work, it is worthwhile reviewing the process by which most enterprise software systems are deployed.

Software Development Life Cycle

The Software Development Life Cycle (also known as the System Development Life Cycle or SDLC) is the process for planning, developing and implementing enterprise software such as traditional Business Intelligence systems. The SDLC is based on the notion that the software application is intended to provide business users with information that they need to perform certain tasks; in the analytical context, to answer certain recurring questions about business performance. [9]

Figure 1: Software Development Life Cycle

The SDLC consists of a set of clearly defined steps which are used by IT professionals to design, build, test and deliver information systems. Typically the process proceeds through each phase within pre-determined time frames and budgets.[10] The nature of the process is sequential and while there are parts of the design process that can be iterative, in general the users first see a running solution many months or quarters after they have stated their requirements.

SDLC Challenges

With business moving at an increasingly rapid pace, one of the challenges of the SDLC is that during the typical 12- to 24-month cycle for delivering functioning analytical applications, business requirements can change more than 50%. The resulting solution may be outdated before it is deployed. In addition, business users are often unable to articulate all of their analytical requirements because, when confronted with a particular business problem, they are not sure which questions to ask. If the business cannot articulate their questions or explain their analytical line of thinking, then IT cannot build the right solution – leaving the business back where they started. This conundrum leaves IT to formulate a solution spec based on their best judgment to enable them to move forward with design, development and deployment. Ultimately an incomplete understanding of business requirements combined with unknown questions and rapidly changing market conditions tends to bring about sub-optimal solutions and a tense relationship between Business and IT.

A new challenge has arisen in the last few years. As a result of datafication, data volume is exploding and data velocity continues to rise, but it is the *variety* of data sources and types that is the biggest challenge. This big data, coming from sources including industrial machines, all sorts of consumer devices, social media services and GPS devices, can provide invaluable insight into trends and future performance, but only if there is a way to include it in analysis.

This situation brings us to ask: How can IT be expected to develop a specification for an analytical solution when data velocity is increasing, when data variety continues to expand, and when some of the potentially most valuable data sources may suddenly appear? While the traditional SDLC process is lengthy and rigid, Data Discovery offers a solution for the complex environment of big data and rapid change.

Stages of the Data Discovery Process

There are five stages of the Data Discovery process: Prototyping, Visualization, Bridging, Replication and Transformation. None of these concepts are by themselves new or unique, but when taken together in the context of Data Discovery, they define a new way to work. The following sections will review each of these concepts.

Prototyping

"No matter your business, no matter your experience, odds are the water's rising. You don't have time, and if you don't act soon you – or your project – will be underwater. There probably isn't time to do things the regular way – or by the book." [11]

At the heart of Data Discovery is the notion of experimentation. Tom Kelley, co-founder of industrial design shop IDEO, calls this process 'prototyping'. "When the project is especially complex, prototyping is a way of making progress when the challenges seem insurmountable ….prototyping is about acting before you've got the answers, about taking chances, stumbling a little, but then making it right." [12]

In the absence of a proven formula or a known way of solving a problem, and under time pressure, prototyping lets people test a hypothesis and explore alternative approaches. With Data Discovery, people are empowered to wander around in the data, try new combinations of sources, filter and refine the analysis and find previously hidden patterns. Or jump to another experimental thread if the first one yields no insight.

Experimentation through Data Discovery takes the form of three main tasks: asking new questions, seeing new patterns and adding new data (Figure 2). These steps comprise a continuous process which is iterative and which flows in any direction, depending upon what the user sees or theorizes at any moment.
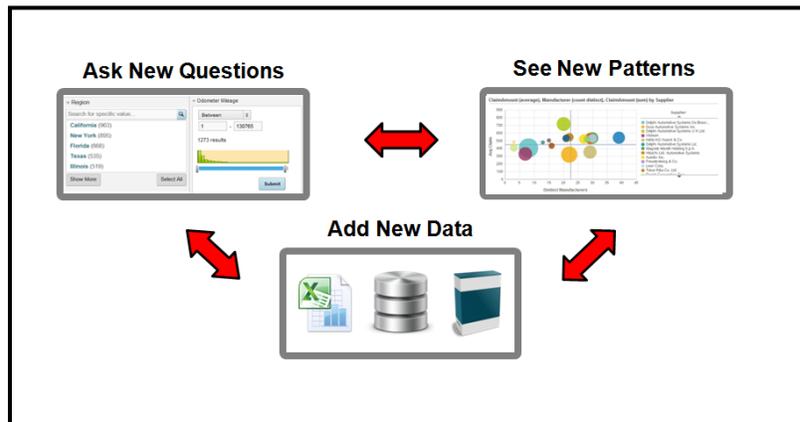

Figure 2: Data Discovery Process
(Source: Oracle)

Visualization

Experimenters want immediate feedback on their hypotheses. It is important for people to see the results of empirical business analysis in real time, to enable them to determine what refined or further search is needed that could lead to deeper understanding of a performance gap or market opportunity. Quick visualization of analysis results is a key component of effective Data Discovery, accomplished for non-technical users through drag and drop dashboards, quickly produced charts, easy to use wizards and consumer style navigation. These techniques accelerate understanding: "…graphics are instruments for reasoning about quantitative information." [13]

For example, at Delphi Automotive: "By combining common analytical techniques like data visualizations and alerts with familiar search features like type-ahead query completion, spell correction and tag clouds, [our] warranty engineers now investigate problems in new ways, asking questions previously thought too difficult to even attempt." [14]

Bridging

"Any CIO's dream is to have a business partner that really embraces technology and wants to do really cool stuff with you. Someone who has a vision, but doesn't come to you and say, 'We want

you to use this product to do this.' They come to you with a problem and they want to work together to figure out how to solve it." [15]

Data Discovery can bring about a more effective working relationship between Business and IT. Instead of a scenario in which Business and IT play separate, traditional provider-versus-user roles, in Data Discovery they combine efforts to jointly explore and learn. IT analysts become more like tour guides to the discovery process, sitting next to their Business partners and stepping through the capabilities of the software, showing them how to add data sources, refine searches and explore new questions that come to instantly to mind. They don't have to worry about extracting requirements, writing complex specs and going through a lengthy development and deployment process. They show the Business what is possible, how to build discovery applications in real time, and how to become self-sufficient in analysis. As data sculptors, IT gains huge efficiencies in bypassing much of the SDLC process and by empowering business users in self service Data Discovery.

Data Discovery tools like Oracle Endeca Information Discovery are intended for direct use by line of business individual contributors and front line managers. As business experts they typically do not have the time or know-how to write sequel queries. They want IT to help them learn how to explore the data and experiment their way to insight. They are able to show their IT counterparts what they are trying to understand. As they become more self sufficient they become self guided data explorers.

As people sit down together to experiment and to explore the data, the entire process becomes more collaborative, interactive and productive. People on both sides can *show* each other what they mean and what they want, build trust and bridge a gap of understanding. Attaining greater business insight at lower cost and higher speed becomes a competitive advantage.

Systematizing

As data explorers become increasingly proficient with their new skills, they may find that in some cases a certain path of investigation is repeated time and time again. Or that it yields insight that is valuable beyond a limited one-time scenario. It may bring to light a question that continually recurs and is useful for tracking or understanding a part of the business.
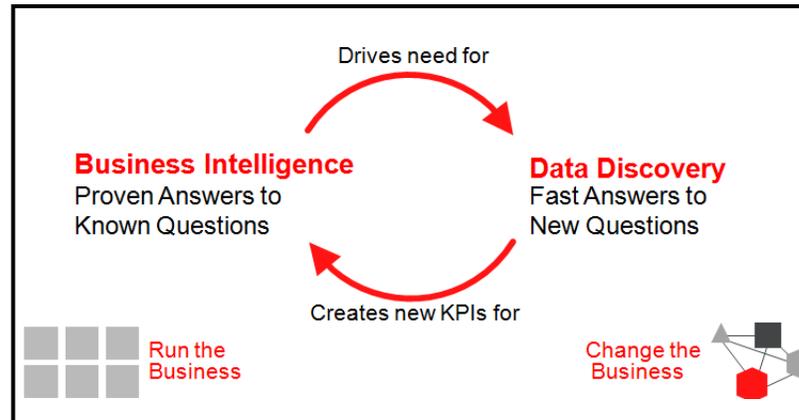
Figure 3: Complementary Interaction of BI and Data Discovery
(Source: Oracle)

In cases where the questions are well known - driven by the metrics and dimensions that are used to describe key business processes – it is worth considering whether to roll the analysis and metrics into the organization's enterprise BI system. Business Intelligence platforms excel in allowing users to perform standardized queries for known questions on standard data sets. Figure 3 shows how Data Discovery and Business Intelligence systems complement each other, with Data Discovery excelling in unknown questions and BI focusing on systematized analysis.

Transformation

Transformation occurs in the minds of users when they realize that they have uncovered valuable new insight through a process of experimentation. The Data Discovery process can help them explore an unlimited number of new avenues to solve problems and find new business opportunities previously hidden in the data. For standard, known questions they will continue using BI systems, but for new questions and problems, they can now explore their way to insight.

When faced with a business analysis problem that had no existing solution at Land O' Lakes, the Endeca team said "Give us a bunch of data, we'll put together a prototype and show you the possibilities." Business users were ecstatic: " 'This is *exactly* (sic) what we're looking for'….from the day we released this (solution), the response was overwhelmingly positive." [15]

Data Discovery is proving its ability to deploy easy exploration on diverse data in weeks, uncover insight that drives dramatic increases in revenue and productivity, and improve the working relationship of Business and IT. It enables a transformation in the world of Business Analytics.

Governance in the Big Data Era

"While much of what we've learned for governing structured, in-house data will be applicable for big data environments, big data also causes us to re-think some of our established data governance procedures." [16] Access to the increasing variety and volume of Big Data raises new questions about how we use and manage that data. How do we govern data that is not produced or managed by the enterprise? What are the guidelines for data quality, retention, security and privacy in an environment of Hadoop clusters and NoSQL databases? As we mash up combinations of data, how do we know whether the new findings actually make sense; where and when to use the insights gleaned from that data?

To illustrate, consider retail chains that sell grocery, pharmacy and general merchandise under one roof. Mashing up consumer purchase data may provide valuable insight for promotions and cross-sell, but prescription-related pharmacy data is HIPPA compliant and subject to restricted use. In some countries, privacy controls restrict the use of business and consumer data for marketing purposes. In financial services and telecom industries, consumers may opt-out of certain databases, meaning that the data can only be used for limited purposes.

Some of the factors that should be considered in this new era are: [16]

- Data stewardship: who should be the business owners of data, particularly that data that is sourced externally?
- Data definitions: how should we establish metadata requirements for new sources of data?
- Data quality: what are policies for acceptable standards of data quality, and who will monitor them?
- Data storage: what are policies for updating, archiving, maintenance and deletion?
- Data security: which data must be classified as private, regulated or confidential, and how will restrictions affect its use?

One approach to Big Data governance is to screen the data before it is consumed in analysis or discovery. This process of filtering may involve tagging data sources, assigning values for attributes such as accuracy, validity and privacy and monitoring then on an ongoing basis. The goals are to improve the value of insight drawn from the data and prevent dissemination of data-driven conclusions that are based on restricted information.

New Roles

New roles will emerge in the era of Data Discovery. IT Business Analysts will do less requirements gathering and solution specification, and more discovery consultation. Instead of translating business requirements into technical specs, analysts who are conversant in both business and technology will become the authors of discovery applications. Line-of-business subject matter experts (SME's) who previously were users of BI applications will become analytic storytellers, applying their understanding of the business as they explore for patterns and insight in the data.

What types of people and skills might be best suited for Data Discovery? People who are multi-dimensional, who can blend science with art, ideas with logical reasoning. People who are adept at discovery – detectives, research scientists, anthropologists – those who look for clues and follow them to establish correlations. Business Analysts who today operate at the boundary between business process and technology, and who can migrate beyond structured roles towards experimentation and exploration.

Other new roles may be needed. The Data Detective is someone who is adept at foreseeing data set relationships that may yield new insights, and who understands the difference between correlation and causation (correlation may be incidental; causation implies cause-and-effect). The Experimenter is the person who tests the validity of new discoveries. The Formalizer is a person who takes validated discoveries and rolls them into the BI infrastructure to provide proven answers to known questions, and raises the level of best practice.

Big Data Leadership

"Crafting and implementing a big-data and advanced-analytics strategy demands much more than serving up data to an external provider…Rather, it's about effecting widespread change in the way a company does its day-to-day business." [17] Managing data in this new world highlights the need for the role of the Chief Data Officer (CDO), potentially reporting to the CIO. As a C-level position the CDO's role is "establishing new mind-sets around the value of data…" [18], providing leadership and guidance around Big Data policies and procedures. It is to help the organization know "…when and where to use the insights gleaned from the data. For example, do employees know when and how to act upon the insights gained from social media sites like Facebook and LinkedIn?" [19]

With greater interaction and integration between the worlds of Business Intelligence and Data Discovery (Figure 3) the concept of an analytical Center of Excellence (COE) may take on a larger and more meaningful role in the organization. It can serve as a mechanism for defining policies and procedures and for providing guidance and leadership around Data Discovery.

In a typical COE there are three principal domains: People (mentioned above), Process and Governance.[20] In Governance, the COE may provide for the oversight and filtering of data that is internal, external and increasingly varied in source and type. In Process, business analysis has begun to shift from analysis of known questions to discovery around new questions, while increasing in speed, depth and insight. Many of the services that the COE enables may continue, but the way in which they are delivered may radically change – from a process of specification to a focus on exploration.

Conclusion

To prepare for the coming age of Data Discovery, senior Business and IT leaders should take time to learn about the potential of this new solution. They should identify the business problems that have been difficult to understand or solve where new sources of data could provide insight - particularly those that are unstructured, external to the organization or absent from an existing data model. They should think about how Data Discovery will change the way that people work, and how new processes and roles can improve business insight and productivity. They should also try to identify the members of their staff who are inclined to think outside the box, and who are open to embracing a new way of thinking and analyzing.

Footnotes

1. Oracle Corporation. "Are Your Making Decisions with Only 20 Percent of Your Data?" Oracle Corporation: Endeca Business Brief, 2012.

2. Oracle Corporation. "Oracle Endeca Information Discovery: A Technical Overview". Redwood Shores, CA: Oracle White Paper, December 2012

3. Oracle Corporation. "Harnessing the Power of Employee Sentiment". Oracle Corporation: Endeca Business Brief, 2012.

4. Oracle Corporation. "Oracle Endeca Information Discovery: A Technical Overview". Redwood Shores, CA: Oracle White Paper, December 2012.

5. Mayer-Schönberger, Viktor and Kenneth Cukier. Big Data: A Revolution that Will Transform How We Live, Work and Think. Boston: Houghton Mifflin Harcourt, 2103. Page 78.

6. Mayer-Schönberger, Viktor and Kenneth Cukier. Big Data: A Revolution that Will Transform How We Live, Work and Think. Boston: Houghton Mifflin Harcourt, 2103. Page 96.

7. Mayer-Schönberger, Viktor and Kenneth Cukier. Big Data: A Revolution that Will Transform How We Live, Work and Think. Boston: Houghton Mifflin Harcourt, 2013. Page 103.

8. Oracle Corporation. "Oracle Endeca Information Discovery: A Technical Overview". Redwood Shores, CA: Oracle White Paper, December 2012.

9. Beynon-Davies, P. Business Information Systems. Palgrave, Basingstoke, 2009.

10. "Systems Development Life-Cycle". Wikipedia, November 2013.
http://en.wikipedia.org/wiki/Systems_development_life-cycle

11. Kelley, Tom. The Art of Innovation. New York: Currency Books / Doubleday, 2001. Page 104.

12. Kelley, Tom. The Art of Innovation. New York: Currency Books / Doubleday, 2001. Pages 106-7.

13. Tufte, Edward R. The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press LLC, 2001. P. 10.

14. Oracle Corporation. "Delphi Improves Warranty Response and Customer Satisfaction with Oracle Endeca Information Discovery". Oracle Corporation: Endeca Business Brief, 2012.

15. DataInformed article: "Land O' Lakes Analytics Platform Sows Sales Growth with Farmers". Wellesley Information Services, June 6, 2013. http://data-informed.com/land-o-lakes-analytics-platform-sows-sales-growth-with-farmers.

16. Villar, Maria C. "Data Governance and Big Data". Information-Management.com, June 26, 2013.
http://www.information-management.com/news/data-governance-and-big-data-10024554-1.html

17. Brown, Brad, David Court and Paul Willmott. "Mobilizing your C-suite for big-data analytics". McKinsey Quarterly, November 2013.
http://www.mckinsey.com/insights/business_technology/mobilizing_your_c_suite_for_big_data_analytics

18. Griffith, Terry. "Just Adding a Chief Data Officer Isn't Enough". HBR Blog Network, December 11, 2013.
http://blogs.hbr.org/2013/12/just-adding-a-chief-data-officer-isnt-enough/?utm_source=Socialflow&utm_medium=Tweet&utm_campaign=Socialflow

19. Schmarzo, Bill. "Data Governance: On the Road to Decision Governance". January 16, 2012.
https://infocus.emc.com/william_schmarzo/data-governance-on-the-road-to-decision-governance/

20. Krakauer, Jake. "The Business Intelligence Competency Center: Enabling Continuous Improvement in Performance Management." Oracle Corporation White Paper, January 2012. http://www.oracle.com/us/products/middleware/bus-int/bicc-white-paper-1-2012-1486911.pdf

ORACLE®

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

Hardware and Software, Engineered to Work Together