Capture, Organize and Analyze Big Data for Research Enterprise

Rakhisah Mat Zin, Oracle Corporation Malaysia
rakhisah.mat-zin@oracle.com

Today the term 'big data' draws a lot of attention, but behind the hype—especially for research organizations—there's a simple story. For decades, traditional research organizations have been collecting and working with data on a daily basis. As computing technology has evolved, so has the ability to gather, aggregate, analyze, and store increasing volumes of data. But even many of the most forward-thinking research technologists underestimated how fast these volumes of data would grow. And the challenge is how to prepare the framework in managing and processing this big data will less difficulties so that the organization can focus on the research or analysis itself.

Key Words: Oracle Big Data, Engineered Systems, Big Data Discovery, R Enterprise

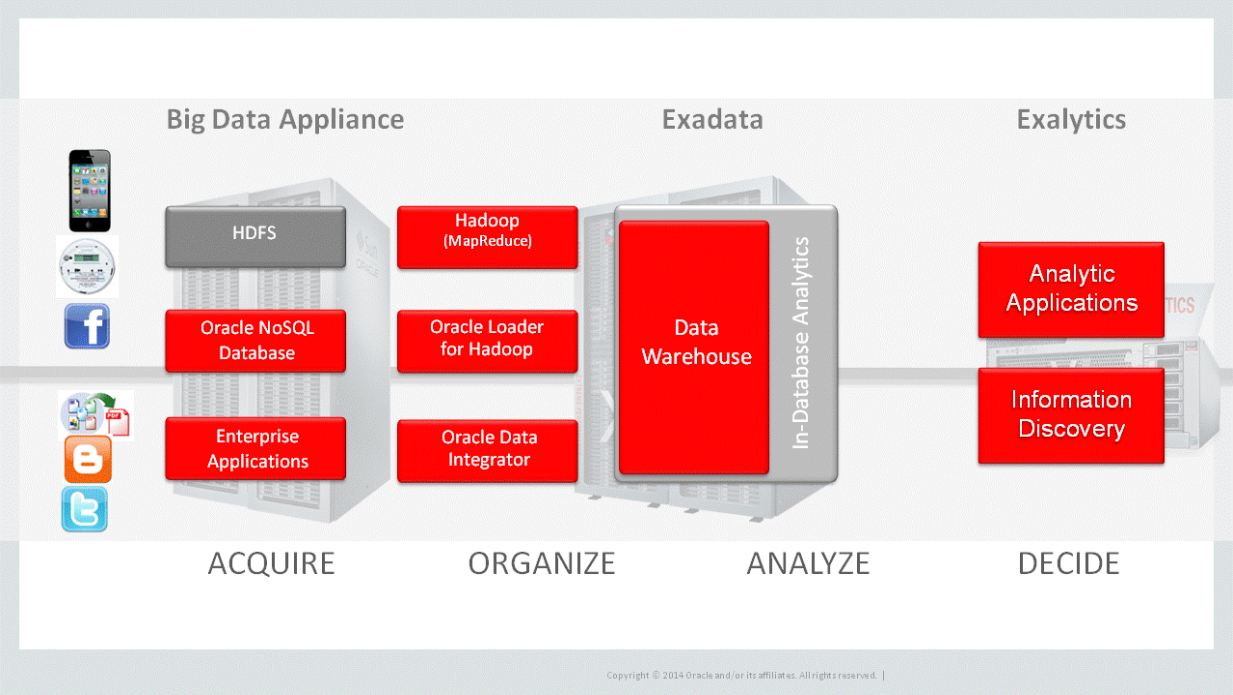
Introduction

In large part, the data deluge—or big data phenomenon— in research has been fueled by the proliferation of unstructured, non-traditional data generated through collaboration tools and social networking sources as well as the global sharing among researchers of observational data, simulation models, and experimental data. In addition, libraries have continued to digitize huge volumes of archived bodies of research that once were available only to a handful of researchers—or not at all.

In addition, the cost of storage, compute power, and capacity has decreased, making it more affordable to aggregate, share, and analyze data in ways that may not have been feasible for many research organizations just a few years ago. And the proliferation of smart phones, GPS and other mobile devices have supported the immediacy of capturing observation and location research data—including large multi-media files.

Oracle Solutions for Big Data Research

Oracle offers the broadest and most integrated portfolio of products to help research organizations acquire, aggregate, organize and analyze data from diverse sources. And Oracle is the first vendor to address the full spectrum of research enterprise big data requirements and is uniquely qualified to combine everything needed to meet your big data challenges—including software and hardware—into one engineered systems.



Key Components of the Research Infrastructure to Support Big Data

The requirements in a big data infrastructure span data acquisition, data organization and data analysis. Oracle's solutions to support these areas are outlined below. For more details, download Oracle's whitepaper on Big Data Whitepaper or visit the information about Oracle Big Data on Oracle.com.

Acquire: Making the most of big data means quickly capturing high volumes of data generated in many different formats. Oracle offers a range of solution which includes Big Data Appliance which bundled with software to support big data processing. As the leader in database technologies, Oracle is developed to handle very high transaction volumes in a distributed environment and support the research community's need for flexible, dynamic data structures.

Organize: A big data research platform needs to process massive quantities of data—filtering, transforming and sorting it before loading it into a data warehouse. Oracle offers a choice of solution for organizing data including Oracle Data Integration together with Big Data Connectors. In addition, Oracle enables end-to-end control of structured and unstructured content, allowing you to manage all your data from application-to-archive efficiently, securely, and cost effectively with the Oracle content management and tiered storage solution designed specifically for research organizations.

Analyze: The infrastructure required for analyzing big data must be able to support deeper analytics such as statistical analysis and data mining on a wider variety of data types stored in diverse systems; scale to extreme data volumes; deliver faster response times; and automate decisions based on analytical models. Oracle offers a portfolio of tools for statistical and advanced analysis. These solutions include the Oracle Exadata and Data Warehousing as well as

developer tools for the application layer such as Oracle Application Express, which allows your researchers to more easily access and analyze data from within their applications.

Being in data management business for more than 30 years, Oracle objective is to reduce data movement to make it easier for customer to process and analyze the data. Oracle Database 12c running on engineered system such as Oracle Exadata allows customer to use in-database implementation of advanced analytics features like data mining and statistical analysis without moving the data out of Oracle Database. This is provided thru the use of Oracle Advanced Analytics function which includes Oracle Data Mining and Oracle R Enterprise. With these technologies, research enterprise will enjoy the benefit of having predictive analytics, data mining, text mining, statistical analysis, advanced numerical computations and interactive graphics inside the database.

Another challenge in implementing big data projects is basically lack of expertise in using new language that comes with new technologies in Hadoop world to access and analyze the data. Oracle has taken one step further in delivering Big Data SQL where we bring the knowledge of SQL to be able to access both structured data from the relational database together with unstructured data from Hadoop world and also NoSQL data. Customers can use existing expertise in using SQL to manage all data sets.

The key benefit of using these technologies is to reduce data movement, preserve data security and maximize performance.

Decide: Many analytical and visualization tools have surfaced in recent years. It is important to understand the similarities and differences of the key capabilities to help select the right tool for the right analysis and user base.

Business Intelligence and Information Discovery solve different problems and creates different type of values. Business Intelligence provides proven answers to known questions. Information Discovery provides fast answers to new questions. The key performance indicators (KPIs), reports, and dashboards produced by the Business Intelligence tools drive the need for exploration and discovery using Information Discovery Platform. Visualization tools that will help users to make decision can be either Oracle Business Intelligence or Oracle Big Data Discovery to support the information discovery needs.

Majority of us are familiar with Business Intelligence since we have been using these tools as part of analytics or data warehouse projects before. This tool normally will work with structured data where the data model already prepared. However, when we move to Big Data world, where the data can be structure, semi-structured and also unstructured, we need to have the capability of analyzing the relationship of this data without creating specific data model and able to discovery new patterns which can help organizations to create new KPIs.

CASE STUDIES

FINANCIAL SECTOR: CAIXABANK – BARCELONA, SPAIN

CaixaBank Maximizes Big Data Business Value and Improves Analytics, Agility, Service, and Organizational Efficiency

Background	CaixaBank is a leading Spanish retail bank and insurer with more than 13.8 million customers, 9,700 ATMs, and 5,300 branches. Centering on delivering customer service through innovation and technology, CaixaBank is one of Europe's leading banks with 4.2 million active internet banking users and 2.6 million mobile bank clients. The bank received the Innovative Spirit Award at the 2014 Global Banking Innovation Awards, selected by Bank Administration Institute and Finacle. CaixaBank was also named Best Retail Bank for Technology Innovation by Euromoney magazine in 2013 and 2014.
Challenges	• Integrate data from bank branches, ATMs, and internet and mobile banking to gain a complete understanding of customers and offer personalized banking solutions—gaining in customer loyalty and business competitiveness • Improve messaging to reach customers more effectively, better informing them of new bank services and products to boost sales • Implement a flexible, adaptable, scalable, and future-proof solution to ensure the bank's competitiveness and ability to adapt to changes in business strategy or banking sector regulations • Input, manage, and analyze massive amounts of diverse data from both external and internal sources, efficiently and effectively extending traditional analysis with latest technologies to discover information, trends, and patterns, regardless of data source or format • Make a complex and expensive data extraction and transformation process agile and flexible, improving time-to-market and enabling the bank to quickly analyze, decide, adapt, and act as the business, its sector, and customers become data-oriented and mobile • Provide employees with access to non-sensitive information to promote creativity and innovation and improve productivity
Solutions	• Consolidated data marts into data pools with Oracle Big Data Appliance, Oracle Exadata, and Oracle Big Data Connectors—integrating massive data from all points of sale and customers' online and mobile profiles—enabling the bank to understand customers, their preferences, and their mood to quickly and flexibly offer them tailored solutions using Oracle Real-Time Decisions, Oracle Advanced Analytics, Oracle R Enterprise, and Oracle Business Intelligence on Oracle Exalytics • Improved click-through rate by 39% and real sells by 50% in less than a month using Oracle Real-Time Decisions, deploying personalized,

	targeted messages at the right moment by leveraging customer data and habits • Gained the ability to find new data patterns, correlations, uses, and transformations using Oracle Advanced Analytics, boosting the bank's agility, flexibility, and time-to-market and enabling it to quickly make decisions and implement changes to respond to business strategies and adapt to evolving banking regulations • Provided a complete, unified view of all internal and external data relevant to the bank's business needs—including institutional reports, information sources, quality assurance data, security data, standardization data, IT applications, business structure data, and analysis results—enabling it to optimize resources, saving time and costs • Enabled authorized users to easily share and access information, providing a structured platform that enables bank employees to submit suggestions and creative solutions and their superiors to follow up and implement changes
Customer Quote	"Oracle Big Data Appliance, Oracle Advanced Analytics, and Oracle Real-Time Decisions enable us to quickly find patterns and correlations in our customers' online interaction with us. We've improved our business agility and flexibility as well as our ability to know and serve our customers, ultimately focusing on creating value for them rather than solving IT issues." – Luis Esteban Grifoll, Chief Data Officer, CaixaBank
Case Study URL	**http://www.oracle.com/us/corporate/customers/customersearch/caixabank-1-big-data-2648217.html**

EDUCATION SECTOR : VALDOSTA STATE UNIVERSITY, USA

Valdosta State University Identifies Key Correlations in Student Data to Improve Student Retention, Progression, and Graduation Rates

Background	Established in 1906, Valdosta State University (VSU) is an American public university and is one of the two regional universities in the University System of Georgia. Valdosta State includes five colleges offering 56 undergraduate degree programs and more than 40 graduate programs and degrees.
Challenges	• Address a 67% one-year student retention rate, which costs the university US$6.5 million in lost revenue per year • Collect and analyze structured and unstructured student data from multiple sources to more quickly and effectively identify at-risk students, engage appropriate faculty and staff, and

	develop targeted programs to improve student retention • Enhance administrative productivity by accelerating response times to data requests from faculty and staff
Solutions	• Implemented Oracle Business Intelligence Enterprise Edition to provide more than 600 faculty and staff members with rich insight into student academic performance • Provided instant, on-demand access to student class loads and grades to enable faculty to make timely decisions on how to facilitate student success, whether the right path is notifying the student's advisor, conducting one-on-one meetings, or implementing tutoring services • Used Oracle Business Intelligence Enterprise Edition's interactive dashboards to improve administrative productivity by more than 500% by enabling administrative staff to fulfill 90% of all data requests from faculty, department heads, deans, and Office of Financial Aid grants staff within 24 hours—compared to the previous average response time of one to three weeks • Implemented Oracle Endeca Information Discovery to combine structured and unstructured student data—like student surveys and ID card usage data—in one system for analysis • Enabled the university to examine granular data to draw previously unknown correlations to make more informed decisions—for example, determining that students who eat breakfast on campus have a 10% higher retention rate • Used Oracle Endeca Information Discovery to discover that freshmen who work on campus have an 85% retention rate, compared to the general freshman population, which has a 55% retention rate—helping spur a decision to invest US$200,000 in student jobs on campus, which will likely save the university US$2 million in retention costs over two years • Streamlined IT responsibilities and cut database administration time in half, enabling database administrators and programmers to focus on core tasks, such as implementing new web applications, exploring new technologies, and building new reports
Customer Quote	"With Oracle Business Intelligence and Oracle Endeca Information Discovery, Valdosta State University has achieved long-term goals in a very short period of time. We have much better insight into student data, helping us to identify at-risk students, promote student success, and improve graduation and retention rates." – Brian Haugabrook, CIO,

	Valdosta State University
Case Study URL	**http://www.oracle.com/us/corporate/customers/customersearch/vsu-1-endeca-ss-2156270.html**

Summary

No matter what type of data your researchers need to capture, access, and analyze and decide; Oracle's industry-leading solutions deliver the capacity, security, and processing speed researchers want. Oracle has all the infrastructure components needed to support big data sharing and interoperability for global, heterogeneous, and increasingly multidisciplinary research environments and ecosystems.