INTEGRATE FOR INSIGHT

Enterprises must learn to understand how best to leverage big data soon, since the amount of data being generated shows no signs of slowing down.

InfoWorld
Custom Solutions Group

Combining big data tools with traditional data management offers enterprises the complete view

Big data continues to be the topic of much discussion and hype, and companies that have pioneered ways to analyze big data and integrate it with traditional data are finding that the benefits are very real.

Big data—information gleaned from nontraditional sources such as blogs, social media, email, sensors, photographs, video footage, etc., and therefore typically unstructured and voluminous—holds the promise of giving enterprises deeper insight into their customers, partners, and business. This data can provide answers to questions they may not have even thought to ask. What's more, companies benefit from a multidimensional view of their business when they add insight from big data to the traditional types of information they collect and analyze. For example, a company that operates a retail Web site can use big data to understand site visitors' activities, such as paths through the site, pages viewed, and comments posted. This knowledge can be combined with purchasing history and stored in a corporate relational database. From this, the company gains a better understanding of customers, and can fine-tune offers to target their interests.

Enterprises must learn to understand how best to leverage big data soon, since the amount of data being generated shows no signs of slowing down. The McKinsey Global Institute estimates that data volume is growing 40 percent per year, and will grow 44-fold between 2009 and 2020. With this explosion in data comes a wealth of new opportunities for companies to improve business processes, new product development, customer service, brand awareness, product revision cycles, and partner networks—all by mining information that's readily available.

"The question CIOs should be asking now is `where do I start, and how?'" says George Lumpkin, vice president of product management for Oracle's data warehousing business. "The time is right to be looking at big data, and many organizations are starting to use these techniques. Those that aren't will find themselves at risk for being left behind."

Web companies are among the early adopters of big data, largely because of the volume of unstructured information that they must deal with on a regular basis. However, even traditional industries such as telecommunications, retail, financial services, and health-

> **"The time is right to be looking at big data, and many organizations are starting to use these techniques. Those that aren't will find themselves at risk for being left behind."**
>
> —George Lumpkin
> Vice President, Product Management
> Oracle

care are launching pilots and testing the waters to see what big data has to offer. While companies are starting to realize the benefits and stand to gain a competitive edge from leveraging big data, they are also faced with new challenges that require different ways of thinking and innovative technology approaches.

THE BIG DATA DIFFERENCE

Big data is like traditional data in many ways: It must be captured, stored, organized, and analyzed, and the results of the analysis need to be integrated into established processes and influence how the business operates. But because big data comes from relatively new types of data sources that previously weren't mined for insight, companies aren't accustomed to collecting information from these sources, nor are they used to dealing with such large volumes of unstructured data. Therefore, much of the information available to enterprises isn't captured or stored for long-term analysis, and opportunities for gaining insight are missed.

"Because of the huge data volumes, many companies do not keep their big data, and thus do not realize any value from their big

THE ORACLE APPROACH

Oracle offers a broad portfolio of products to help enterprises acquire, manage, and integrate big data with existing information, with the goal of achieving a complete view of business in the fastest, most reliable, and cost effective way.

The Oracle Big Data Appliance is an engineered system of hardware and software designed to help enterprises derive maximum value from their big data strategies. It combines optimized hardware with a comprehensive software stack featuring specialized solutions developed by Oracle to deliver a complete, easy-to-deploy offering for acquiring, organizing and analyzing big data, with enterprise-class performance, availability, supportability, and security.

The Oracle Big Data Appliance incorporates Cloudera's Distribution, including Apache Hadoop with Cloudera Manager, plus an open source distribution of R, all running on Oracle Linux. The Oracle Big Data Appliance comes in a full rack configuration of 18 Oracle Sun servers and scales by connecting multiple racks together via an InfiniBand network, enabling it to acquire, organize, and analyze extreme data volumes.

The Oracle Big Data Appliance offers the following benefits:

- ▶▶ Rapid provisioning of a highly-available and scalable system for managing massive amounts of data
- ▶▶ A high-performance platform for acquiring, organizing, and analyzing big data in Hadoop and using R on raw-data sources

- ▶▶ Control of IT costs by pre-integrating all hardware and software components into a single big data solution that complements enterprise data warehouses

Oracle Big Data Connectors is an optimized software suite to help enterprises integrate data stored in Hadoop or Oracle NoSQL Databases with Oracle Database 11g. It enables very fast data movements between these two environments using Oracle Loader for Hadoop and Oracle Direct Connector for Hadoop Distributed File System (HDFS), while Oracle Data Integrator Application Adapter for Hadoop and Oracle R Connector for Hadoop provide non-Hadoop experts with easier access to HDFS data and MapReduce functionality.

Oracle Exalytics In-Memory Machine is purpose-built to deliver the fastest performance for business intelligence (BI) and planning applications. It is designed to provide real-time, speed-of-thought visual analysis, and enable new types of analytic applications so organizations can make decisions faster in the context of rapidly shifting business conditions, while broadening user adoption of BI though introduction of interactive visualization capabilities. Organizations can extend BI initiatives beyond reporting and dashboards to modeling, planning, forecasting, and predictive analytics.

These offerings, along with Oracle Exadata Database Machine and Oracle Database 11g, create a complete set of technologies for leveraging and integrating big data, and help enterprises quickly and efficiently turn information into insight.

data. Think about Web sites generating logs: there really hasn't been a cost-effective way to capture and store that data before. So it's just being discarded," says Oracle's Lumpkin.

New technologies are emerging that enable organizations to get their arms around these vast quantities of unstructured data, making the prospect of gaining insight both feasible and cost-effective. For example, Hadoop is an open-source platform for consolidating, combining, and transforming large data volumes. MapReduce is a programming framework to support processing large data sets on distributed nodes to generate aggregated results. Key enablers for analyzing big data are tools for statistical and advanced analysis. These tools must be able to work with distributed data to perform analysis regardless of where the data resides, to scale as big data volumes grow, to deliver response times driven by changes in behavior, and to automate decisions based on analytical models.

However, managing these vast quantities of data is only half the battle. Companies that want to truly benefit from big data must also integrate these new types of information with traditional corporate data, and fit the insight they glean into their existing business processes and operations.

To get the complete picture, enterprises are advised to integrate Hadoop with traditional database environments, adding big-data sources to the existing corporate data the organization has built up over the years. Companies that have learned to truly leverage big data understand that they must analyze their new sources of information within the context of the bigger picture; in other words, integrating big data with traditional data to gain a 360-degree view of the extended enterprise.

> **Companies that have learned to truly leverage big data understand that they must analyze their new sources of information within the context of the bigger picture; in other words, integrating big data with traditional data to gain a 360-degree view of the extended enterprise.**

Not only does this approach offer a more complete understanding of the business, it also builds upon existing IT architectures instead of replacing them.

"Big data has a huge potential impact, but organizations can't sweep aside existing architectures—for any organization this has to be evolutionary," says Lumpkin.

With this 360-degree view of their business, enterprises have the insight they need to improve processes and gain a competitive edge. This insight has implications that go far beyond technology to organizational structures, hierarchies, and a company's ability to change, however, so enterprises are advised to take a phased approach to leveraging big data. Initial projects should be small in scope: identify one set of desired data, capture it, explore new data-management techniques, and determine integration points with existing data. Starting with pilot projects and building on successes will help enterprises realize the benefits of leveraging big data with minimal disruption to the business.

Emerging technologies that address big data enable enterprises to analyze new types of information that hadn't been feasible to analyze before. By adding this information to the mix, enterprises can leverage the best of structured and unstructured data to find new solutions to business challenges, become closer to their customers, and significantly increase employee productivity through streamlined business processes. Oracle, for one, is leading the way with innovations—such as the Oracle Big Data Appliance, Oracle Big Data Connectors and Oracle Exalytics In-Memory Machine—that promise big payoff from big data (see sidebar). It's now up to companies to jump in and begin the process of making the promise reality. ▌▌

George Lumpkin
VICE PRESIDENT
PRODUCT MANAGEMENT
ORACLE
Lumpkin is vice president of product management for Oracle's data warehousing group.

FOR MORE INFORMATION:
please visit **www.oracle.com**

Multidimensional Data

Integrating big data into corporate information architecture gives companies new insight.

Big data is the hot trend in IT today, but leveraging new types of data to turn information into insight presents many challenges. Oracle's George Lumpkin discusses the benefits and hurdles.

Why is leveraging big data so important to business these days?

Analysis of big data—including new types of data that haven't been analyzed before—provides a deeper level of insight into what customers are thinking and how the business operates. The potential payoffs are improving customer retention, selling individual customers more products, and producing items with higher quality and lower rates of return. Studies show that, with proper use, big data can really improve the bottom line to make an impact on overall profitability.

What are the challenges to leveraging big data?

If you look at how systems are used today, we have data warehouses built using relational technology like Oracle Database and Oracle Exadata, which give insight into how the business is running and how to improve it. Big data in some ways is a natural evolution of that. What's different in big data is the new data sources; new types of data not previously captured, stored, or analyzed. Very large volumes of data are acquired very rapidly, and may not be neatly structured, which can make storing and analyzing it a challenge.

Where are most companies on the big-data learning curve?

Today you have all of this data being generated and very often it's not stored for long-term analysis because there really hasn't been a cost-effective way to capture and store it, so companies therefore aren't getting value from it. However, today, there are new techniques that are making it more cost effective and easier for everyone.

The other thing is a lot of data didn't exist before. For example, if you look at the rise of mobile devices, there's a tremendous wealth of information generated about hundreds of millions of consumers. And now organizations are using Hadoop to analyze this data and then integrating it with their traditional enterprise data in their data warehouse to get a full view of their customers. These organizations see Hadoop as complementary to what's already been built for analysis.

So what do big data solutions look like today?

It's really about using technologies like Hadoop for managing new sources of data, but not forgetting everything that's been built up over the past 20 to 25 years. Hadoop becomes an extension of that, adding big data sources into your overall information architecture.

What advantages does Oracle offer for companies looking to analyze big data?

Oracle offers the broadest and most integrated enterprise-ready big data platform, which includes the Oracle Big Data Appliance, Oracle Big Data Connectors, Oracle Exadata and Oracle Exalytics In-Memory Machine. By integrating and optimizing the technologies needed to acquire, organize, analyze and decide on big data, Oracle has made it very easy to jumpstart and maintain big data projects. Also, because Oracle is the single point of contact for support and upgrades, problem resolution is much faster and customers do not incur the cost and effort of tracking and building diverse open source projects.

ORACLE®

InfoWorld
Custom Solutions Group