



The disruptive power of big data

How big data analytics is transforming business





Table of contents

3	Executive overview: The big data revolution
4	The big data imperative
4	Why big data is different
4	New data from new sources—and lots of it
5	Web 2.0 and social media
5	Rich media
6	Sensors
7	Big data analytics—a new approach for big data
9	The big data challenges
9	Challenges for IT
10	Challenges for business leaders
11	HP Enterprise Big Data Platform
11	Vertica - A solution for big data analytics
12	Learn more



Executive overview: The big data revolution

Digital data is doubling in size every two years. That alone is an interesting fact. But what's really compelling is that businesses are harvesting and using this data to improve market knowledge, enhance competitiveness, and transform their operations and even their business models. It's a disruptive change for business. And like the Internet and the computer itself, enterprises that learn to use big data and information management for business advantage will thrive in the big data era; those that do not will find themselves outpaced by more nimble competitors and will risk extinction.

Businesses have always sought to mine actionable information from data. But capitalizing on the vast potential value of big data requires a new approach. Traditional data warehouses and business intelligence (BI) tools don't provide answers fast enough. Big data analytics—tools and processes created specifically to turn big data and analytics into business insight—allow business leaders to develop a conversational relationship with data. It's an approach where answers suggest better questions and refinement leads to insight.

This paper explains the big data revolution—what it is, what opportunities it creates, and what challenges exist. It explains why having a conversation with your data is a fundamentally different way to extract business insight and how big data analytics—and the Hewlett Packard Enterprise Vertica platform—help you do just that. It's a challenge businesses must meet if they hope to thrive in the big data era.



The big data imperative

Why big data is different

Big data typically refers to a volume of data so great that traditional IT systems and technologies can no longer store, manage, and process it. But this isn't just a case of data growth outstripping technology growth. Big data embodies fundamental differences that necessitate new approaches and technologies. Doug Laney of Gartner Group identified the trend toward big data analytics and characterized it using the now-famous three **Vs**, which have been widely adopted and adapted:

- Volume—the amount of data being created
- Velocity—the speed at which data is created and must be processed
- Variety—the array of different data sources and formats

Several factors have prompted the escalation in the volume, velocity, and variety of data.

New data from new sources—and lots of it

As businesses have expanded—and expanded onto the Internet—the volume of business transactions has grown. Walmart is the world's largest retailer, but smaller businesses have also experienced a boom in transaction volumes as a result of Internet initiatives, and they have expanded the data that they collect. Going beyond processing orders, they have implemented new systems to track marketing programs, customer data, warranty claims, complaints, and other information. Much of this data is siloed—stored, managed, and processed by disparate systems. And much big data value is derived from simply bringing together data from many information management sources to achieve a 360-degree view of customers, products, and business operations.

Web-based systems capture not only sales transactions but also the click streams that lead to the sale. Merchandisers can see the other products that purchasers viewed and considered and how they reacted to onsite advertising and promotions. The ubiquity of mobile phones—and accompanying Internet-based ordering and support—has resulted in an explosion in call data records, customer data, and usage information available to telecommunications carriers. Automation in the healthcare industry and the consolidation of care offered through large providers like HMOs, insurers, and government programs have created massive amounts of data. The financial industry is casting a broad net to capture data that helps identify trends, assess risk, and manage assets.

Web 2.0 and social media

Web 2.0 goes beyond letting people access information on the Internet. It has revolutionized how people interact and collaborate with each other, their business partners, distribution networks, and supply chains. People contribute their own content, and applications get richer as more people use them. Social media is the outgrowth of Web 2.0. Facebook now has more than 1 billion active subscribers,¹ and it's estimated that they share almost 3.3 million individual pieces of content every minute. In that same minute, Twitter users send 347,000 tweets.²

Practically all businesses have implemented Web 2.0 concepts, and many now have social media marketing teams to determine how to use social media to reach customers and turn social media content into business value. Product reviews, user communities, forums, blogs, and the comments that accompany them all contain valuable business insight. Most companies set up Facebook pages to connect with users, and they seek Twitter followers for their brands and even key personnel. This lets them broadcast information to consumers and capture the reaction to announcements and other changes.

What's more, much of this data is publicly available. Twitter, for example, provides an application program interface (API) that allows third parties to access public tweets. (All tweets are public by default.) And they have contributed their entire archive of public tweets to the U.S. Library of Congress. The availability of this kind of unstructured text data has led to new techniques and tools for "sentiment analysis"—the ability to detect opinion and emotion contained within textual information.

Rich media

Multimedia files such as video, audio, and images are proliferating at an unprecedented rate. Partly driven by the growth of smartphones and the increasing presence of security cameras, the growth of rich media represents a significant source of insight.

The value of rich media has long been known to law enforcement agencies, which now survey a massive amount of video feeds, including videos captured on smartphones and shared via social media, to solve crimes. Enterprises are also seeing dramatic growth in rich media: Video and audio use has catapulted in recent years with the increase in online meetings, and a tremendous amount of valuable information is being captured in multimedia training sessions, video conferences, webcasts, demos, and human resources material. Daily call-center conversations, which inform your customers' impression of your brand, are another source of invaluable information.

The value of rich media is also apparent in consumer trends. According to a [report by Cisco](#), video traffic currently accounts for 78 percent of consumed Internet bandwidth in the United States, and the report projects that by 2018, 84 percent of U.S. Internet traffic will be consumed by video. Those numbers are probably no surprise to anyone who is familiar with YouTube or who uses a news aggregator. Video is a primary communication medium across the Web. So are photos, especially with the rise of smartphones. Photos are constantly being uploaded to social media sites, and the information contained in these images reveals valuable clues about the consumers who post them. Their hobbies, clothing preferences, food choices, political leanings, peer groups, charitable interests, and much more are revealed in these images.

¹ Facebook Newsroom, newsroom.fb.com/company-info/

² Official About Twitter page, Feb 11, 2015, about.twitter.com/company

Sensors

Devices are beginning to compete with humans in the amount of information they generate. For the purposes of this paper, we'll call any such device a sensor—something that senses, captures, and reports information.

For example, cell phones “sense” and report their locations using global positioning systems (GPS), and businesses use that information in location-based services to deliver more targeted advertising and promotions. Retailers are experimenting with radio frequency identification (RFID) sensors and other solutions to track the location of customers within stores so they can target promotions and gauge response to merchandising initiatives. The healthcare industry seeks to improve and extend care and reduce cost through remote patient monitoring. The automotive industry is embedding sensors in vehicles and insurance companies are creating usage-based insurance models based on that sensor data. Utilities are deploying smart meters to identify usage trends for things like water and electricity, and helping consumers reduce their utility costs by consuming these commodities during low-demand periods. McKinsey Global Institute reports that more than 30 million networked sensors are in use in the transportation, automotive, industrial, utilities, and retail sectors, and the number is growing by 30 percent annually.³

Corporate big data

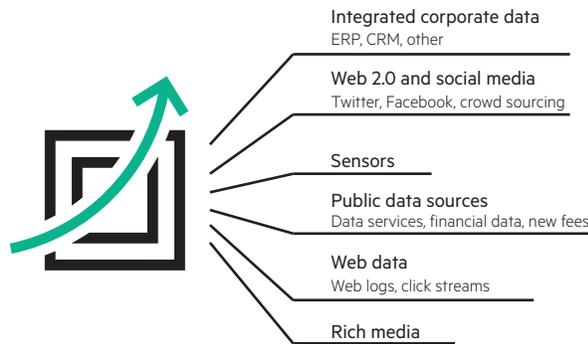


Figure 1. Big data comes from new sources in unprecedented volumes.

³ “Big Data: The next frontier for innovation, competition, and productivity,” The McKinsey Global Institute, June 2012.

Big data analytics—a new approach for big data

All the data sources mentioned above, and more, are available to businesses. But what does the data mean? How can business organizations use big data to make operations more efficient, improve competitiveness, and enhance revenue? Why do traditional data warehouses and BI tools fall short in the big data era? And what must businesses do to not just survive but thrive.

Business leaders seek insight that enables them to act and then assess the results of their actions—to gain new knowledge into their markets, products, and operations. That's the traditional role of BI. But traditional systems aren't designed to handle the volume, velocity, and variety of big data. The bottom line for business leaders is that these systems produce answers to their questions hours or days after they ask them. If the answer doesn't deliver the insight needed, they can ask a better question, but it takes hours or days to get that answer, too. It's like trying to conduct a job interview using letters and post cards.

What thriving businesses need is an iterative, more conversational approach to real-time analytics. When systems produce answers in minutes or seconds, business analysts can immediately apply that answer to frame better, more meaningful questions, which, in turn, produce more insightful answers. They can ask, "What if?" And each answer suggests better questions until they achieve the desired insight. The processes, technologies, and skills needed to do that are collectively called business analytics, and we refer to business analytics scaled up to handle the volume, velocity, and variety of big data as big data analytics.

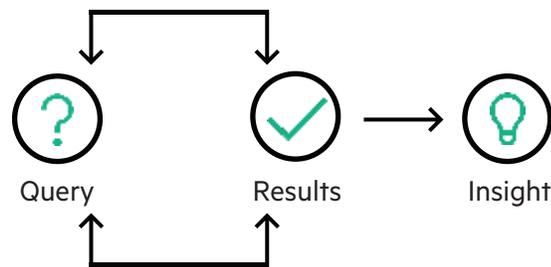


Figure 2. A conversational approach to business analytics



Thriving businesses use big data analytics to:

- **Understand market reaction and brand perception.** Businesses can assess consumer reaction to product announcements, pricing changes, policy changes, and other moves. They can gauge the effect of promotions, advertising campaigns, and publicity programs. And they can compare sentiment associated with their brands to those of competitors.
- **Identify key buying factors.** Analysis of website click streams, product reviews, product registration information, and public reaction to products can help merchandisers and product planners understand what buyers seek and why they buy what they buy.
- **Segment populations to customize actions.** When marketers can understand the attributes of different segments of the market and how members of each segment react to products, advertising, and promotions, they can tune marketing actions to achieve the best results within each segment.
- **Enable experimentation.** When business leaders can quickly assess the effect of changes in products, marketing, and operations, they can perform experiments to help assess the value of changes and determine the best course.
- **Predict outcomes.** Insight into the result of previous changes coupled with better understanding of market reactions lets businesses better predict the results of changes like price increases, advertising, and promotions.
- **Enable new business models.** In many cases, information derived from big data analytics has value outside the enterprise that collected it. Information services that aggregate and analyze big data have grown up in industries like healthcare, real estate, finance, retail, and others.

Each of these represents new insight opportunities that have only recently become available to businesses. And each is now becoming a business basic for the 21st century. That's why big data is a disruption to—not just an evolution of—business operations. Businesses that master big data analytics and information management will move ahead of less-nimble competitors.

Let's look at the challenges to see why traditional approaches are preventing business leaders from making the leap to big data analytics.

The big data challenges

BI techniques and technologies have been in widespread use since the 1990s. To better understand where our customers were in big data capture and application, Hewlett-Packard commissioned Coleman Parkes, a market research company, to survey senior business and IT executives in enterprise-level companies in October, 2012. The results:

- 34 percent say half their information is unused.
- 35 percent believe they are not effective at accessing enterprise information.
- 50 percent do not have an effective information strategy in place.
- 98 percent say they cannot deliver the right information at the right time.

Collecting, storing, and analyzing big data present new challenges, both for IT organizations and for business leaders. And traditional data warehouses and BI systems have not helped businesses extract the value from big data. Why?

Challenges for IT

Let's look at the technical side first. Most databases in use today are based on relational database management system (RDBMS) technology that was developed in the 1980s. They are designed to process and store business transactions, so their structure and access methods are focused on transactions. Database schemas are highly structured and not very flexible. Relational databases store data associated with individual transactions together in rows, so adding, finding, updating, and deleting transactions is fast and easy—just what you need in a transaction-processing database.

The concept of the data warehouse also evolved in the 1980s. Data warehouses were simply databases into which data from multiple sources was consolidated for the purpose of querying and reporting. Data warehouses do not need to support transaction processing, so the technologies evolved to handle larger dataset sizes and were optimized for data retrieval rather than adding, updating, and deleting transactional data. Some commercial products achieved the necessary storage by distributing data over multiple systems, and many implementations resulted in terabytes of data storage. But even these systems fall short when confronted with the volume, velocity, and variety of big data. They fail to enable the conversational approach to data required by big data analytics. Here's why.

Traditional databases and even data warehouses don't easily scale to the hundreds of terabytes or even petabytes needed for many big data applications. Large, multi-system implementations are expensive to acquire and maintain. Data is usually not compressed, so huge amounts of storage and I/O bandwidth are needed to load, store, and retrieve data. In some cases, the velocity of incoming data is such that there are not enough hours in the day to load it into the database. Data is still stored in tables by row, so access to a single data element through many rows—a common operation in business analytics—requires retrieving practically all of the data in a dataset to extract the small amount actually needed. That strains I/O bandwidth and extends processing time. Queries produce answers in hours rather than the seconds or minutes needed for iterative business analytics. As a result, systems are expensive, and they fail to enable the near-real-time conversational approach that big data analytics requires.



Further, traditional systems don't easily accommodate the variety of data types and formats found in big data. It's estimated that 85 percent of big data is unstructured—comprised of free-form text, audio, and video rather than neatly organized into the recognizable fields handled by RDBMSs. New tools and techniques—like sentiment analysis—are used to extract meaningful information from unstructured data so results can be stored in databases and analyzed. But the formats of resulting data are less predictable, much more varied, and subject to change during iterative, real-time analytics. This requires frequent changes to RDBMS structure and to processes that load data into them.

For IT, it means the iterative approach to extracting business insight from big data requires new approaches, tools, and skills.

Challenges for business leaders

Big data is not just a technical challenge. Gaining and applying business insight require that business leaders also adopt new ways of thinking and working.

Successful leaders in data-exploiting organizations become more familiar with the sources of data available to them. Rather than asking IT what information is available in the database, these leaders view information at the level of their industry. They explore how it can be acquired and project what insights might be extracted from it. They challenge IT to acquire the necessary data, tools, and skills. They partner with them to develop and apply the conversational, iterative approach needed by big data analytics. And they team business analysts and IT data specialists to mine the business value it contains.

Even more important, successful leaders apply insight to decision making. In the past, decision makers examined the limited data available, and relied on experience and intuition to bridge the gap between data and insight. But J. Paul Getty said, "In times of rapid change, experience could be your worst enemy." And this is certainly a time of rapid change. Business leaders can now achieve insight based on analysis of real data. As they apply it in the decision-making process, they gain new experience to be applied to new business problems—a prescription for thriving in the 21st century.



HP Enterprise Big Data Platform

Vertica - A solution for big data analytics

HPE Vertica is the industry's first comprehensive, scalable, open, and secure platform for big data. HPE Vertica, a massively scalable analytical database platform, is custom-built for real-time analytics on petabyte-sized datasets. It supports standard SQL, Python and R-based analytics, and offers support for all leading BI and ETL vendors.

Our analytical platform is available to those who want to install it on-premises, on their Hadoop backbone or in the cloud. Users can leverage the same analytical engine no matter where their data sits or where it came from. Vertica delivers the most advanced SQL analytics as a standards-based relational database with full support for SQL, JDBC, and ODBC. This allows users to preserve years of investment and training in these technologies because all SQL programming tools and languages work seamlessly together. HP Vertica also includes a wide array of built-in analytical functions, including geospatial, time series, pattern matching, and more, and supports all leading BI and visualization tools.



Learn more

Are you one of the 98 percent of businesses that say they cannot deliver the right data at the right time? Do you believe the insight available in big data can help your business thrive? Then learn more about the HPE Vertica platform. HPE has helped more than 600 businesses and government agencies use big data to improve operations, gain competitive advantage, and find new revenue opportunities.

Test drive the HPE Haven platform at havenondemand.com.

HPE also offers HPE Vertica Community Edition software, a free version of HPE Vertica Enterprise Edition limited to one terabyte of data and three nodes. Sign up for HPE Vertica Community Edition at vertica.com/community.

Get connected with HPE
hp.com/go/getconnected

Sign up for updates

★ Rate this document