

Business white paper

The disruptive power of Big Data

How Big Data analytics can transform your business



Table of contents

3	Executive overview: the Big Data revolution
4	The Big Data imperative
4	Why Big Data is different
6	Big Data analytics—a new approach for Big Data
7	The Big Data challenges
7	Challenges for IT
7	Challenges for business leaders
8	HP Vertica Analytics Platform
8	A solution for Big Data analytics
8	Learn more

Executive overview: the Big Data revolution

Harvard Business Review says 2.5 exabytes (billion gigabytes) of data are created every day.¹ It's no wonder. In that same day, Facebook users share 1 billion pieces of content, and Twitter users generate more than 200 million tweets. Two million users access the Internet to search and buy, leaving behind click streams, comments, and product reviews. Corporate data has also blossomed to almost unmanageable volumes. Walmart processes more than 1 million customer transactions every hour.² And smaller enterprises roll out new applications in every part of their business, each of which collects vast amounts of new information. So much data is available that the information technology (IT) industry has spawned a new term to describe it: "Big Data."

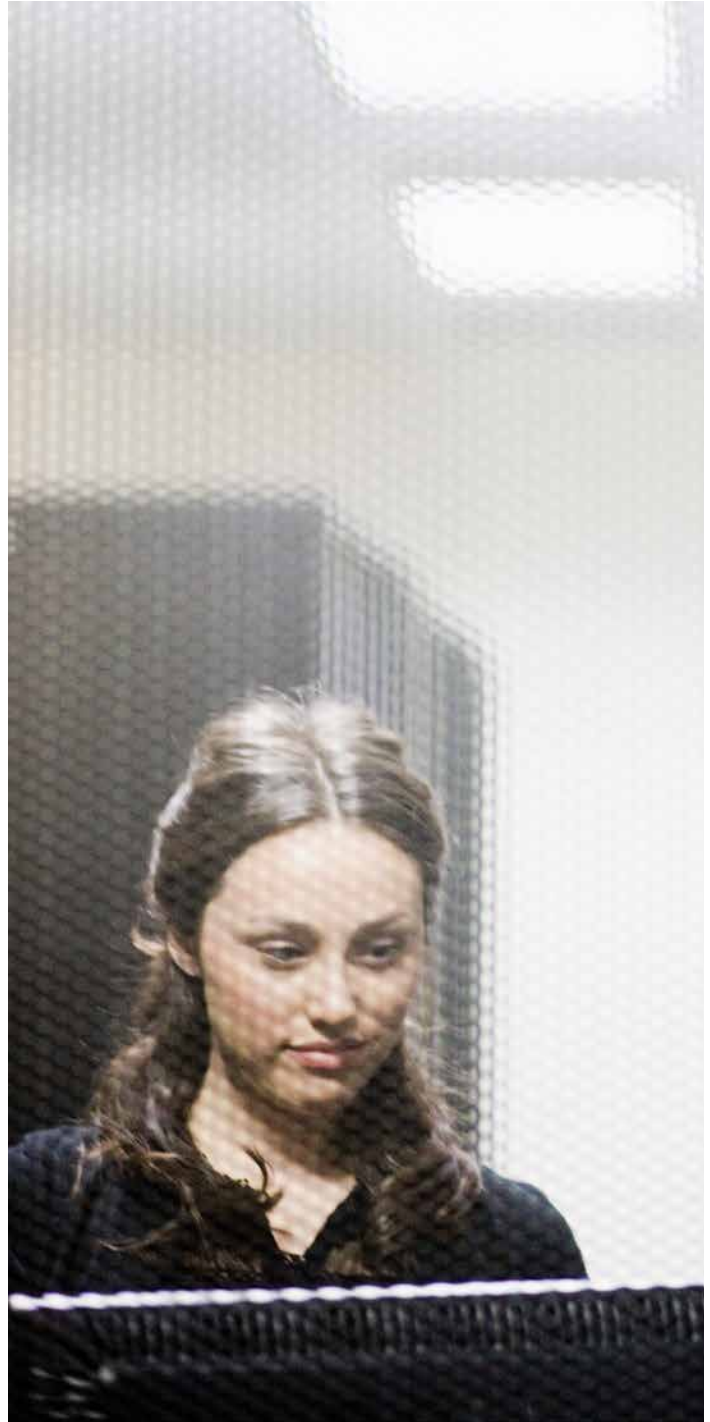
The fact that the volume of data in the world is exploding is interesting. What's compelling, though, is that businesses are harvesting and using this data to improve market knowledge, enhance competitiveness, and transform their operations and even their business models. It's a disruptive change for business. And like the Internet and the computer itself, enterprises that learn to use Big Data for business advantage will thrive in the Big Data era. Those that do not will find themselves outpaced by more nimble competitors and 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 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 HP Vertica Analytics Platform—help you do just that. It's a challenge thrivers must meet.

¹ "Big Data: The Management Revolution," Andrew McAfee and Erik Brynjolfsson, *The Magazine*, Harvard Business Review, October, 1012, <http://hbr.org/2012/10/big-data-the-management-revolution/ar/1>.

² "Data, data everywhere," *The Economist*, February 25, 2010, economist.com/node/15557443?story_id=15557443.



The Big Data imperative

Why Big Data is different

The most common definition of Big Data is data in such volumes 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 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. McKinsey Global Institute estimates that in most U.S. business sectors, companies with more than 1000 employees store more than 235 terabytes of data.³ 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 sources to achieve a 360 degree view of customers, products, and business operations.

Web-based systems not only capture sales transactions, but the click streams that lead to the sale. Merchandisers can see the other products that purchasers viewed and considered and how they reacted to on-site advertising and promotions. The near 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 health care 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, and with 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,⁴ and it's estimated that they share almost 700,000 individual pieces of content every minute. In that same minute, Twitter users send 140,000 tweets.⁵

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 allow the consumer to generate content that may 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.

³ "Are you ready for the era of 'Big Data'?", Brad Brown, Michael Chui, James Manyika; McKinsey Quarterly, October, 2011.

⁴ Facebook Newsroom, newsroom.fb.com/content/default.aspx?NewsAreaId=22.

⁵ Twitter Official Blog, June 30, 2011, <http://blog.twitter.com/search?q=200+million+tweets>.

Sensors

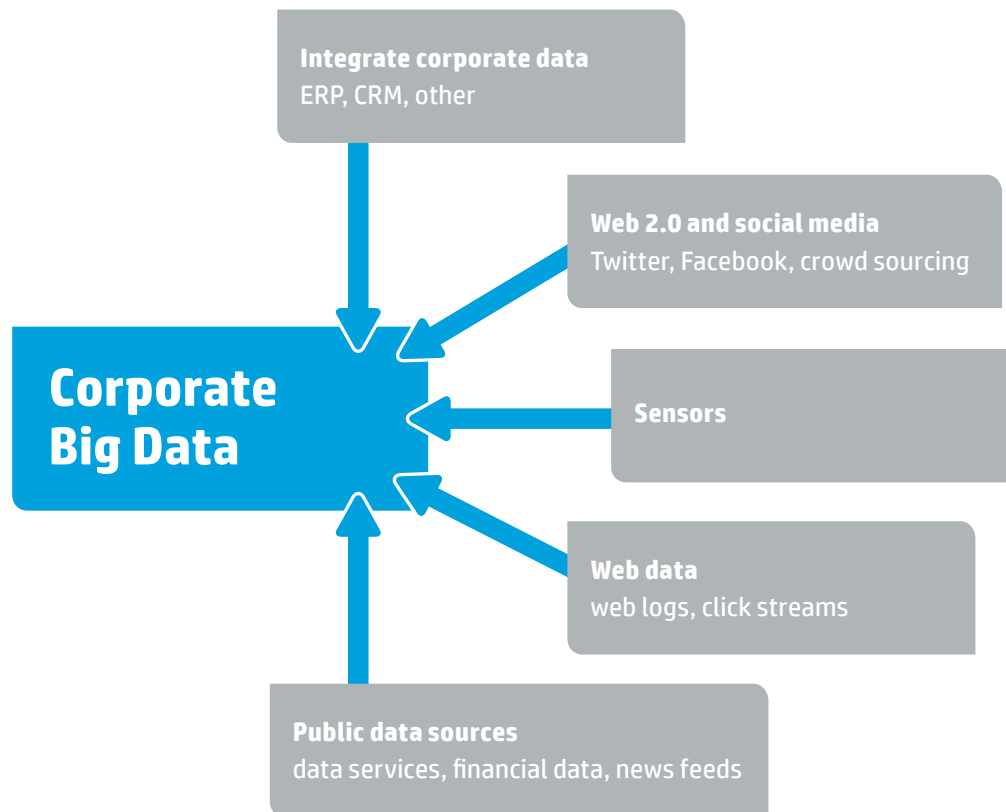
The amount of information generated by devices rather than the conscious acts of humans is growing steadily. For our purposes, we'll call any such device a sensor.

Cell phones that report their locations using Global Positioning System (GPS) are an example. Businesses are using that information in location-based services to deliver more targeted advertising and promotions. Retailers are experimenting with radio frequency identification (RFID) sensors and other approaches to track the location of customers within stores, so they can target promotions and gauge response to merchandising initiatives. The health care 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 reduce costs and improve their response to usage trends. 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% annually.⁶

⁶ "Big Data: The next frontier for innovation, competition, and productivity," The McKinsey Global Institute, June, 2012.

Figure 1: Big Data comes from new sources in unprecedented volumes.



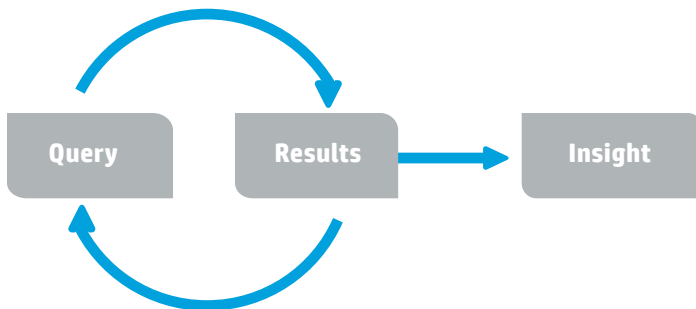
Big Data analytics—a new approach for Big Data

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 be among the thrivers?

Business leaders seek insight that enables them to act and assess the results of their actions—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. We’ll highlight some of the technical reasons later, but 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 thrivers need is an iterative, more conversational approach to 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?” Each answer suggests better questions until they achieve the desired insight. The processes, technologies, and skills needed to do that is called business analytics. 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



Thrivers 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 has value outside the enterprise that collected it. Information services that aggregate and analyze Big Data have grown up in industries like health care, real estate, finance, retail, and others.

Each of these represents new insight opportunities that have been unavailable to businesses in any significant way in the past. 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 will move ahead of less nimble competitors.

Now let’s look at the challenges to see why traditional approaches don’t let business leaders make that required leap.

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, we commissioned Coleman Parkes, a market research company, to survey senior business and IT executives in enterprise-level companies in October 2012. The results:

- 34% say half their information is unused.
- 35% believe they are not effective at accessing enterprise information.
- 50% do not have an effective information strategy in place.
- 98% 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 transaction data. Some commercial products achieved the necessary storage by distributing data over multiple systems, and many implementations reached into the terabytes of stored data. 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. We have seen cases where 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 conversational approach to Big Data analytics needed.

Further, traditional systems don't easily accommodate the variety of data types and formats found in Big Data. It's estimated that 85% 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—let us 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 analytics. This requires frequent changes to RDBMS structure and the 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 requires that business leaders also adopt new ways of thinking and working.

Successful leaders we have known 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, they 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, insightful leaders apply gained 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 Vertica Analytics Platform

A solution for Big Data analytics

HP Vertica Analytics Platform is a new kind of data analytics platform designed from the ground up for business analytics at the scale of Big Data. Compared to traditional databases and data warehouses, it drives down the cost of capturing, storing, and analyzing data. And it produces answers 50 to 1000 times faster to enable the iterative, conversational analytics approach needed.

Here's why:

- HP Vertica Analytics Platform compresses data to reduce storage costs and speed access by up to 90%.
- It stores data by columns rather than rows and caches data in memory to make analytic queries 50 to 1000 times faster.
- It uses massively parallel processing (MPP) to spread huge data volumes over any hardware, including low-cost commodity servers.
- It uses data replication, failover, and recovery to achieve automatic high availability.
- It includes a pre-packaged, in-database analytics library to handle complex analytics and a robust development framework and support for the R statistical programming language to enable analysts to create user-defined analytics inside the database.
- It dynamically integrates with and complements Hadoop to move large sets of structured, semi-structured, and unstructured data back and forth between systems for data exploration and fast analytics.

Taken together, HP Vertica Analytics Platform means better, faster business insight at less cost.

Learn more

Are you one of the 98% 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 HP Vertica Analytics Platform. We've 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 HP Vertica Analytics Platform at vertica.com/evaluate.

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

Get connected with HP: hp.com/go/getconnected.

Get connected

hp.com/go/getconnected

Get the insider view on tech trends, support alerts, and HP solutions.



Share with colleagues

© Copyright 2012 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

4AA4-4527ENW, Created November 2012

