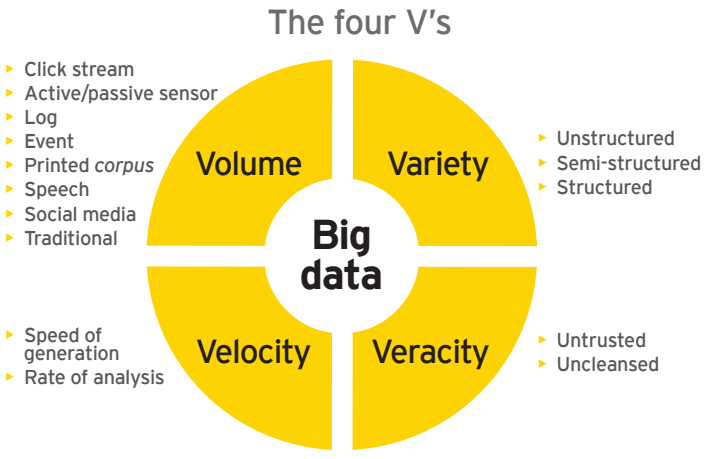
What is big data?

Big data refers to the dynamic, large and disparate volumes of data being created by people, tools and machines; it requires new, innovative and scalable technology to collect, host and analytically process the vast amount of data gathered in order to derive real-time business insights that relate to consumers, risk, profit, performance, productivity management and enhanced shareholder value.

Big data includes information garnered from social media, data from internet-enabled devices (including smartphones and tablets), machine data, video and voice recordings, and the continued preservation and logging of structured and unstructured data. It is typically characterized by the four "V's":

- **Volume:** the amount of data being created is vast compared to traditional data sources
- **Variety:** data comes from different sources and is being created by machines as well as people
- **Velocity:** data is being generated extremely fast – a process that never stops, even while we sleep
- **Veracity:** big data is sourced from many different places, as a result you need to test the veracity/quality of the data



Evolving technology has brought data analysis out of IT backrooms, and extended the potential of using data-driven results into every facet of an organization. However, while advances in software and hardware have enabled the age of big data, technology is not the only consideration. Companies need to take a holistic view that recognizes that success is built upon the integration of people, process, technology and data; this means being able to incorporate data into their business routines, their strategy and their daily operations.

Organizations must understand what insights they need in order to make good strategic and operational decisions. The first part of the challenge is sorting through all of the available data to identify trends and correlations that will drive beneficial changes in business behavior. The next step is enriching this organizational information with that from sources outside the enterprise; this will include familiar big data sources, such as those created and stored online.

In a business environment that constantly and rapidly changes, future prediction becomes more important than the simple visualization of historical or current perspectives. For effective future prediction, data analysis using statistical and predictive modeling techniques may be applied to enhance and support the organization's business strategy. The collection and aggregation of big data, and other information from outside the enterprise, enables the business to develop their own analytic capacity and capability, which for many years has only been available to a few larger organizations.

The impact of big data

To understand the impact of how data has transformed our daily lives, look no further than how the movie rental experience has changed. When movies were rented from independent neighborhood stores, the rental agent would base their recommendations on which movies the customer said they liked and a large amount of their own opinion.

Today, movie rental companies and content delivery services can utilize a vast array of data points to generate recommendations. By analyzing what was viewed, when, on what device (and even whether the content was fast forwarded, rewound or paused), as well as user activities such as internet searches, and browsing and scrolling within a webpage, recommendations can be tailored for millions of customers in real time andapproximately 75% of views at a leading provider are now driven by these recommendations.

Technology megatrends

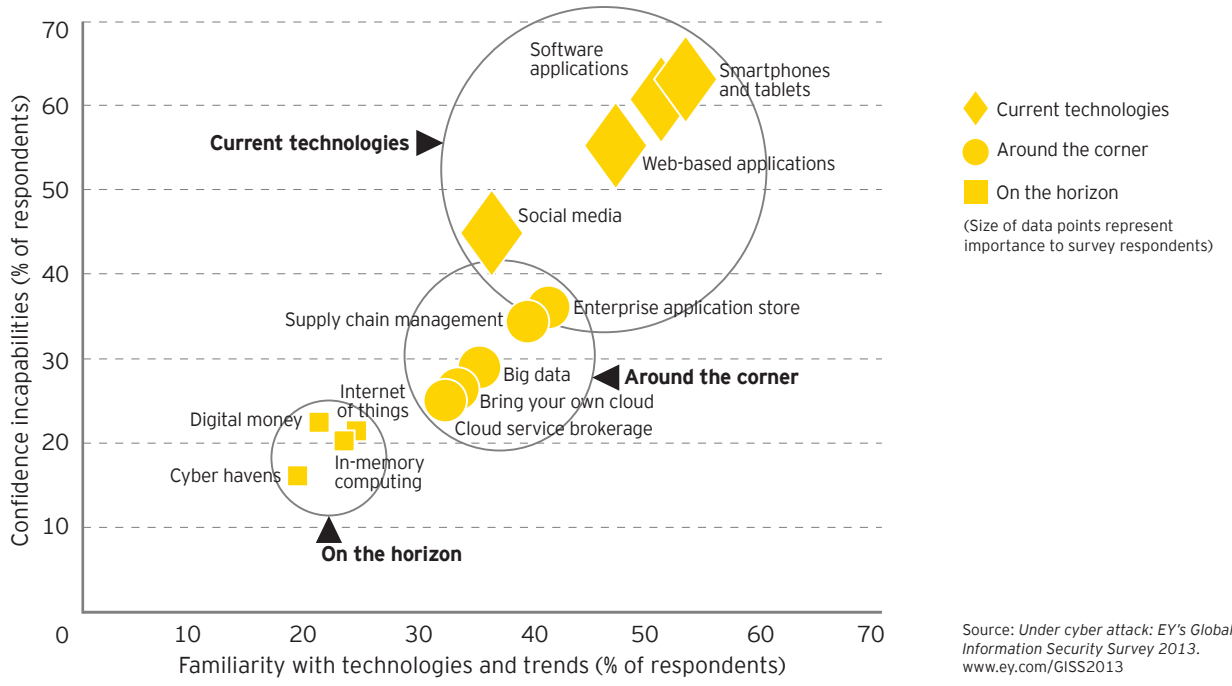
Big data is generating an intense amount of attention among businesses, media and even consumers, along with analytics, cloud-based technologies, digital channels and data visualization. These are all part of the current diverse ecosystem created by the technology megatrends. Some even herald the potential transformative power of the current trends as rivaling that of the internet. Yet, as in the early days of the internet, there is uncertainty about just what big data is, its potential benefits and the associated risks.

EY's 2013 Global Information Security Survey results indicate that while adoption and use of big data is not yet widespread, there is growing confidence and familiarity with the technology. Respondents ranked big data technologies as being "around the corner" (i.e., those that have been on organizations' radar for a period of time but may not yet be implemented or widely adopted) as average in terms of level of importance, familiarity and confidence in their capabilities to address related cyber risks. Organizations typically view these technologies as offering opportunities to improve their performance and create competitive advantage. This is where familiarity and confidence in capabilities needs to increase today, as the importance of these technologies is likely to grow significantly in the near future.

The term Big Data has become a major theme of the technology media, but it has also increasingly made its way into many compliance, internal audit and fraud risk management-related discussions. In EY's *Global Forensic Data Analytics Survey 2014*, 72% of respondents believe that emerging big data technologies can play a key role in fraud prevention and detection. Yet only 7% of respondents were aware of any specific big data technologies, and only 2% were actually using them.

Forensic data analytics (FDA) technologies are available to help companies keep pace with increasing data volumes, as well as business and regulatory complexities; examples can include real-time analytical processing engines that make rapid business decisions, such as stopping a potentially improper payment or business transaction, or leveraging anti-fraud/anti-corruption monitoring controls that integrate data visualization, statistical analysis and text mining. Yet despite their availability, many companies have not scaled up their data usage to take advantage of these effective tools, and may be missing important fraud prevention and detection opportunities by not mining larger data sets to more robustly monitor business activities.

Emerging technologies and trends



Source: *Under cyber attack: EY's Global Information Security Survey 2013*. www.ey.com/GISS2013

Big data life cycle

Creation

Certain types of data have long been able to be captured, but this data has rarely been used effectively until now (e.g., the location of a person at any point in time, the number of steps a person takes every day, a real-time history of credit card purchases). New technology such as advanced sensors and customized software can now record this information for analysis.

Changes in the way we communicate (e.g., social media vs. telephone vs. text/SMS vs. email vs. letter) have also increased our ability to investigate areas such as consumer sentiment. Social media increases the speed at which data is generated; for example, a product launch that is discussed live on a popular social networking site can generate a buzz in real-time and allow companies to gauge public reaction even before the launch event is over.

Processing

Extremely large volumes of data have traditionally not been captured and processed for various reasons, most notably because the cost to do so was far greater than the value of insights companies could derive from its analysis. However, multiple factors and new technologies have lowered the cost and technology barrier for effective data processing, allowing companies of all sizes, to be able to unlock the value contained in different data sources. For instance, it is difficult for conventional relational databases to handle unstructured data, so software frameworks like Hadoop(R), for distributed storage and parallel processing of large datasets have been introduced to process non-structured data at high speed; making it easier to perform a more comprehensive analysis of big data.

Many organizations are looking to the cloud to provide a storage solution that is agile and enables unparalleled scalability; however, these organizations need to ensure the governance and risk management practices on their cloud are appropriate for the type of information being collected. Cloud computing enables companies to use prebuilt big data solutions, or quickly build and deploy a powerful array of servers, without the substantial costs involved in owning physical hardware.

Output

Although it is now easier and cheaper to capture, store and process data, it is not useful unless the information is relevant; it must also be readily available to the right people who need the appropriate input in order to make insightful decisions leading to successful outcomes.

There are three key enablers:

- **Mobile** – established mobile networks have allowed for easier distribution of information in real-time
- **Visual/interactive** – technologies have brought the ability to review large and complex data sets into the realm of the average business user
- **Human resource** – there is a new breed of employees with the knowledge to handle the complexities of big data and with the ability to simplify the output for daily use

Resources and processes

An important factor in being able to achieve big data success is having knowledgeable and competent resources. This extends beyond the so-called data scientists who have deep knowledge and experience in handling, analyzing and reporting on big data sets. While these skillsets are indeed in high demand, success requires more than having a handful of specialists on the workforce.

While governments continue to call for the training of "data scientists," companies are taking advantage of fundamental skillsets that already exist within their organization. Employees with the curiosity to ask the right questions and the ability to synthesize and leverage new data points quickly are well suited to lead the big data revolution. In reality, *they are the revolution,* but they must be supported with business processes that place value on gathering and using data, and that integrate data-driven decision making.

Big data and analytics

Big data poses both opportunities and challenges for businesses. In order to extract value from big data, it must be processed and analyzed in a timely manner, and the results need to be available in such a way as to be able to effect positive change or influence business decisions. The effectiveness also relies on an organization having the right combination of people, process and technology.

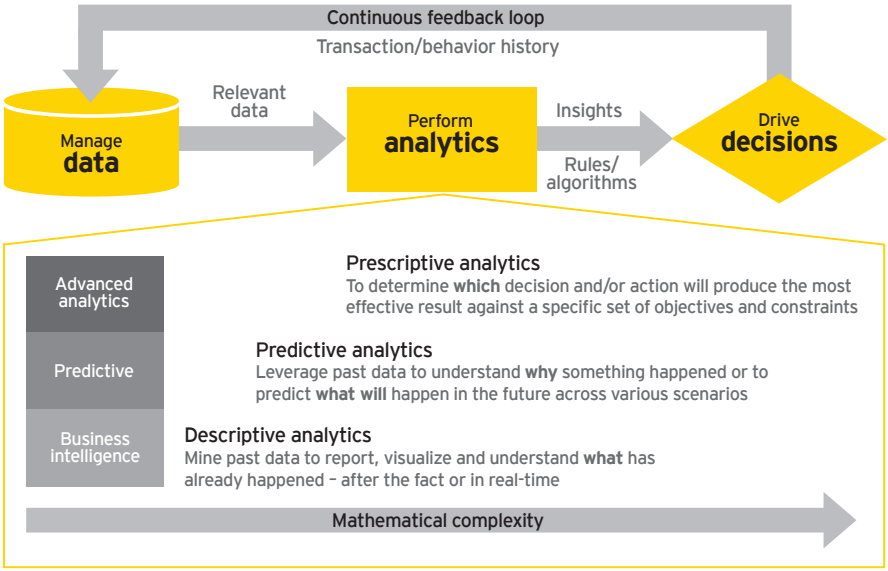
By pure definition, analytics is the discovery and communication of meaningful patterns in data – but for business, analytics should be viewed as the extensive use of data, statistical and quantitative analysis, using explanatory and predictive models to drive fact-based business management decisions and actions.

Analytics helps to optimize key processes, functions and roles. It can be leveraged to aggregate both internal and external data. It enables organizations to meet stakeholder reporting demands, manage massive data volumes, create market advantages, manage risk, improve controls and, ultimately, enhance organizational performance by turning information into intelligence.

EY analytics value chain

The goal is to use analytics to improve the **efficiency** and **effectiveness** of every **decision** and/or **action**.

1. Begin with leveraging leading tools and techniques to manage and extract relevant data from big data sources.
2. Applications of analytics can range from historical reporting, through to real-time decision support for organizations based on future predictions.
3. Use the insight generated by the analysis to drive change.



Analytics can identify innovative opportunities in key processes, functions and roles. It creates a catalyst for innovation and change — and by challenging the status quo, it can help to create new possibilities for the business and its customers. Sophisticated techniques can allow companies to discover root causes, analyze microsegments of their markets, transform processes and make accurate predictions about future events or customers' propensity to buy, churn or engage.

It is no longer enough for companies to simply understand current process or operations with a view on improving what already exists, when there is now the capacity to question if a process is relevant to the business, or whether there is a new way of solving a particular issue. The key driver for innovation within organizations is to constantly challenge existing practices rather than consistently accept the same.

Most organizations have complex and fragmented architecture landscapes that make the cohesive collation and dissemination of data difficult. New analytic solutions are playing an important role in enabling an effective Intelligent Enterprise (IE). An IE helps to create a single view across your organization by utilizing a combination of standard reporting and data visualization:

- Data from multiple source systems is cleansed, normalized and collated
- External feeds can be gathered from the latest research, best practice guidelines, benchmarks and other online repositories
- Use of enhanced visualization techniques, benchmarking indexes and dashboards can inform management and consumers via smartphones, laptops, tablets, etc., in-house or remotely

All companies need to start thinking about collecting and using relevant big data. Data-driven decisions can reduce inefficiency between the business, legal and IT, optimize existing information assets and address disconnects between different functions of an organization. However, it is worth noting that the best data and the most advanced analytical tools and techniques mean nothing if they are not being leveraged by people who are asking the right questions. Big data, emerging storage technology platforms and the latest analytical algorithms are enablers to business success — not a guarantee of it.

Big data can be a powerful way to identify opportunities, but when combined with traditional organizational information the volumes of data collected can be vast and traditional storage methods can be prohibitively expensive and do not necessarily scale effectively.