



knowsquare .

ANTONIO GARCÍA SANSIGRE

14 DE MAYO DE 2014

BIG DATA AT WORK

RESEÑA DEL LIBRO DE TOM DAVENPORT

knowsquare .

Privado y Confidencial

Prohibida su Distribución sin Autorización Expresa del Autor
y Know Square S.L.

Sobre el autor

Thomas Davenport es profesor de gestión y sistemas de información en Babson College. Con anterioridad, ha sido profesor de Harvard Business School, Dartmouth Tuck, y University of Chicago. Es doctor por Harvard en ciencias sociales.

Reseña del libro

According to a recent report, big data market is projected to grow at an annual rate of 40.5% to \$48.3 billion from \$6.3 billion last year. North America will remain the leading contributor to the segment, commanding 55% of the market share¹.

Big data refers to data that is too big to fit on a single server, too unstructured to fit into a row-and-column database, or too continuously flowing to fit into a static data warehouse. It's most difficult aspect is not the size, though, but the fact that it is unstructured.

The first thing to be done with that data is to convert it into insights, innovations. You first need to impose structure on big data. From a business perspective, you can use analytics for competitive advantage. The term "Big Data" started to take off in the fourth quarter of 2010, and originally it seemed a recycled marketing buzzword for data mining. In large, established businesses, it wasn't uncommon for managers to share skepticism and distrust of the hype. But there are real differences between conventional analytics and big data.

	Big data	Traditional analytics
Type of data	Unstructured formats	Formatted in rows and columns
Volume of data	100 terabytes to petabytes	Tens of terabytes or less
Flow of data	Constant flow of data	Static pool of data
Analysis methods	Machine learning	Hypothesis-based
Primary purpose	Data-based products	Internal decision support and services

Only 23 percent said their organizations had a strategy for big data. It can be defined in terms of three Vs (volume, variety, and velocity).

¹ Transparency Market Research titled "Big Data Market – Global Scenario, Trends, Industry Analysis, Size, Share and Forecast, 2012- 2018"

The word is still overused, so it would be more constructive to say “We’re going to analyze video data at our ATMs and branches to better understand customer relationships” rather than “We’re embarking on a big data initiative”.

The idea of analyzing data to make sense of what’s happening in our businesses has been with us for a long time (in corporations since at least 1954, when UPS started an analytics group). From “executive support”, “online analytical processing”, “business intelligence”, “analytics”, “data mining” and now “big data projects” (BDP):

Terminology for using and analyzing data

Term	Time frame	Specific meaning
Decision support	1970–1985	Use of data analysis to support decision making
Executive support	1980–1990	Focus on data analysis for decisions by senior executives
Online analytical processing (OLAP)	1990–2000	Software for analyzing multidimensional data tables
Business intelligence	1989–2005	Tools to support data-driven decisions, with emphasis on reporting
Analytics	2005–2010	Focus on statistical and mathematical analysis for decisions
Big data	2010–present	Focus on very large, unstructured, fast-moving data

These projects must also be connected to the growth of sensors, i.e., the Internet of the Things. The number of networked devices overtook the global population of humans in 2011. Analysts estimate that fifty billion sensors will be connected to the internet by 2025. For instance, farmers are using sensors inside the cow’s stomachs to measure their temperature and detect their sicknesses. Human beings are also being increasingly “sensored”. The earliest popular application of personal analytics I know of was the Nike+/iPod combination released in 2006. Now there are a variety of providers, including JawBone, MyZeo, WakeMate, BodyMedia, and Fitbit.

These BDP need a special kind of professionals: data scientists. Unlike conventional analysts, they are more facile with data, more experimental, and more product-focused. companies. For example, Reid Hoffman, the cofounder and chairman of LinkedIn, made his data scientists a line product team for the company, and they have developed such products as People You May Know, Groups You May Like, Jobs You May Be Interested In, Who’s Viewed My Profile, and several others. Netflix created the well-known Netflix Prize for the data science team that could optimize the company’s movie recommendations for customers.

At the Pentagon, Carey Schwartz, leader of the Pentagon's "Data to Decisions" strategic initiative, is looking to big data technology to make analysts (video and otherwise) a hundred times more productive. He emphasizes the need for "analytics that work," which he describes as analytics that are reliable and robust and capable of being automated.

Another big difference from traditional ERP approach is that BDP consists of agile approaches, in which relatively little time is spent specifying a system up front and more emphasis is placed on creating small deliverables quickly. Peter Drucker commented in 1998 that most information systems were focused on internal accounting data and "have aggravated what all along has been management's degenerative tendency, especially in the big corporations: to focus inward on costs and efforts, rather than outward on opportunities, changes, and threats". Now, with big data, the external focus is beginning to exist.

There are three classes of value: cost reductions, decision improvements, and improvements in products and services.

Many big data enthusiasts argue that there is more value from adding new sources of data to a model than to refining the model itself. For example, Anand Rajaram, who works at WalMartLabs and teaches at Stanford, ran a bit of a natural experiment in one of his Stanford classes along the lines of the Netflix Prize. One of the groups in Rajaram's classes used the data that Netflix provided and applied very sophisticated algorithms to it. Another group supplemented the data (illegally, according to the rules of the competition) with movie genre data from the Internet Movie Database. The latter group's predictions were far more successful. As Peter Norvig, Google's director of research, puts it: "We don't have better algorithms. We just have more data". The trend is so strong that General Electric GE has set up a center in the San Francisco Bay area to address software and big data issues and is hiring lots of data scientists to do so. GE makes a lot of its money these days on servicing industrial products, so if it had proprietary data on how these products are performing and when they might break down, it could promise customers more effective operations and optimize the cost and effectiveness of service operations. Imagine that you were buying a turbine for a power plant. You have the choice of one with big data output, continuous performance monitoring, and optimized service only when it's needed. The other turbine doesn't have those features. Which would you buy, and which might you be willing to pay a bit extra for?

Many organizations have begun to build centrally coordinated analytical strategies and groups. If big data resides in silos and pockets across organizations, it will be very difficult to pull it together to understand and act on business opportunities. The early results from big companies suggest that big data will be combined with existing data and analytics groups, but that could change over the next few years.

However, there is already evidence that customers don't want more targeting—and that they are deeply suspicious of what companies do with their data. It is the "creepiness factor" for customers that might limit the expansion of Big Data.

Transparency vow

El autor de este documento no conoce al autor.

© Antonio García Sansigre

© Know Square S.L.