magpi

BIG DATA AND BUSINESS

What Is Big Data, How Did It Happen, and What Does It Mean for Business?

Table of Contents

What Does Big Data Mean? **3**
How Did Big Data Happen? **4**
Storage 4
Connectivity 5
Computing 6
Big Data Use in Business **8**
Target Knows if You Are Pregnant 8
Mint: Unified View of Financial Datasets 9
Uber Uses Big Data to Improve Transport 9
IBM and NIH Working Together with Big Data 10
Conclusion **11**

What Does Big Data Mean?

There are almost as many definitions of big data as there are people writing about it:

> "data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges"

> "data sets so large and complex that it becomes difficult to process using on-hand data management tools or traditional data processing applications"

> "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze"

No matter what the definition used, the key elements of big data are:

- increasing amounts of data being collected
- increasing connection of datasets
- increasing innovation to manage this data

Of course, there has always been a tremendous amount of information in the world because there has always been a tremendous amount of activity in the world. It is only as we have become better at collecting this information and recording it (thereby rendering it as "data"), communicating about it, and connecting it to other data that the term "big data" could really be considered appropriate.

Nonetheless, the idea that collected data is outstripping our ability to manage it — or at least forcing us to develop new methodologies for managing it — is not a new one. In fact, as early as 1880, it was noted that using the information technologies available at the time it would take 7 years to process the census data. This made it a certainty, given the growing population, that the 1890 census would not be analyzable within the ten year period before the 1900 census began. This prediction led the US government to hire Herman Hollerith to develop a punched card tabulation system to speed up the process.[1]

Hollerith was living in a great age of invention, however, and his tabulating machine was only one of many developments that have brought us to our current stage.

The idea that collected data is outstripping our ability to manage it — or at least forcing us to develop new methodologies for managing it — is not a new one.

[1] https://en.wikipedia.org/wiki/1880_United_States_Census
Hollerith's company, the Tabulating Machine Company, was later joined with several others to create what would become International Business Machines (IBM) in 1924.

How Did Big Data Happen?

Big data can be considered to be the result in large part of enormous advances in three interconnected fields:

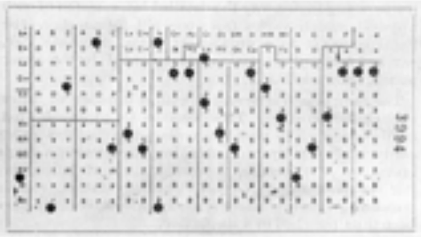
- Storage
- Connectivity
- Computing

It is possible to choose some representative advances — and inventors — in each of these fields, while at the same time emphasizing that these are just that: representative. For every person included, many are necessarily left out.

Storage

Let's begin with the development of modern data storage systems. As mentioned briefly above, in the 1880s the US and other governments were catalyzing development of tabulation systems for the purpose of managing the biggest datasets of that day. Those systems, though, were based on punched paper cards like those used by Herman Hollerith. The standard that developed was a punched card with 80 columns that stored about 70 bytes of data.[2]

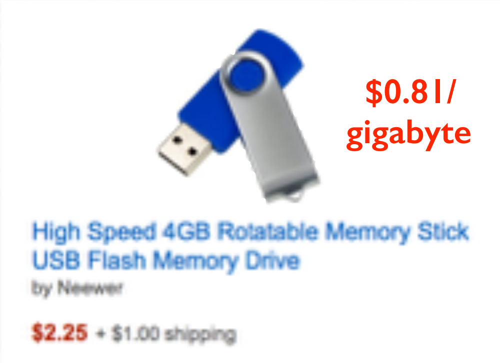
Use of punched card systems by large businesses and the government spurred interest in more "data dense" means of data storage that might be more easily secured (much of the 1890 census data, the first to be tabulated using automated means, was destroyed by fire in 1921).



Hollerith punched card, 1895

In 1898, a Danish engineer named Valdemar Poulsen demonstrated a method of storing sound by means of magnetic wire, and later tape. This technology was adapted for storage of tabulation data, and by the end of the 1940s magnetic storage was the predominant mode of computer storage, and continues to be one of the dominant digital storage methods used today.

According to Parkinson's Law, "work expands so as to fill the time available for its completion", and it has been noted that data storage has its corollary: data expands to fill the space available for storage. This has certainly been the case in the last 4 decades: it is estimated that from the dawn of man until the 20th century a total of 5 exabytes of data (1 exabyte = 1 billion gigabytes) had ever been recorded worldwide, but that we are now recording that same amount



High Speed 4GB Rotatable Memory Stick USB Flash Memory Drive

by Neewer

$2.25 + $1.00 shipping

[2] Note: Hollerith's punchcard above comprises ~150k bytes of data – the equivalent of more than 2000 cards worth of information.

every few days. Luckily, advances in magnetic storage technology, along with other tech advances, have led to a dramatic plunge in storage costs. In fact, the price of one gigabyte of computer storage has fallen from approximately $700,000 U.S. dollars back in 1980, to less than one single dollar today!

Connectivity

Martin Cooper invented the cellular telephone in 1973 while working for Motorola. The first commercial model, called the DynaTac 8000, was made available in 1984. It weighed just under 2 pounds (825g), took 10 hours to charge, offered 30 minutes of talk time, and cost the equivalent of about $10,000.

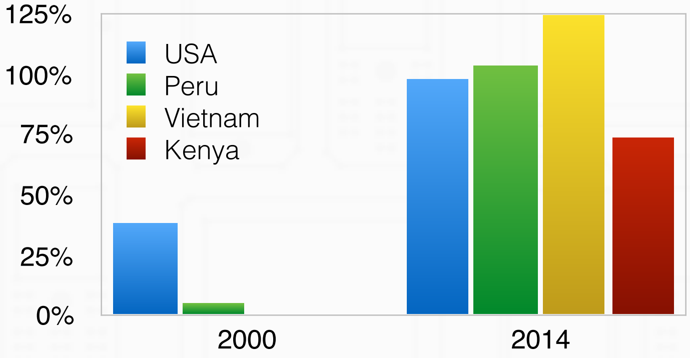
It took more than 20 years for miniaturization technology, cellular networks, economies of scale, and other related developments to create the near-universal current use of cellular phone technology.

As seen in the graph at right, as late as 2000 there was little or no cell phone penetration in poorer countries like Kenya and Vietnam, but currently it is estimated that more than 50% of the population in even the poorest countries has a mobile phone, or use of one.[3]

Economists from the Deloitte consultancy have estimated[4] that each this mobile utilization has tremendous effects on economic growth:

- A doubling of mobile data use leads to a growth in the GDP per capita growth rate of 0.5 percentage points
- Countries characterized by a higher level of data usage per 3G connection have seen an increase in their GDP per capita growth of up to 1.4 percentage points.
- A 10% rise in 3G penetration increases GDP per capita growth by 0.15 percentage points. In developing markets, a 10% expansion in mobile penetration increases productivity in the long run by 4.2 percentage points.

Cell Phone Penetration in Selected Countries, 2000-2014



[3] Note that “penetration” displayed on the graph is measured by the number of SIM cards divided by the population, which overestimates the percentage of the population owning their own mobile phone (since many SIM cards are not yet assigned, and many wealthier individuals may have more than one SIM card). Data: http://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx

[4] Report for the GSM Association: “The impact of mobile on economic growth". http://www2.deloitte.com/uk/en/pages/technology-media-and-telecommunications/articles/impact-of-mobile-telephony-on-economic-growth.html

Computing

As anyone with a smartphone knows, in the last decade Martin Cooper's mobile invention has merged inextricably with the "computer". Compare the IBM PC XT, the dominant business computing platform of the mid-1980s, with the Alcatel One Touch, one of the *least* capable and *least* expensive phones at the time of this writing.

The $10 Alcatel is fifty times faster, with about thirty-thousand times more memory.

And a built-in FM radio and flashlight.

In fact, an iPhone of today is far more powerful than the multi-million-dollar Cray supercomputers of just thirty years ago. Just as amazing as the remarkable increase in power has been the diminishing size of components, and the dramatic decrease in the price – and the fact that making telephone calls is a fast-diminishing use of such a "telephone," and usage shifts more to the computing side and away from telephony.[5]

1985

IBM
PC XT
128 KB
4.8 MHz
$3000

2015

Alcatel
One Touch
4GB
250 MHz
$10

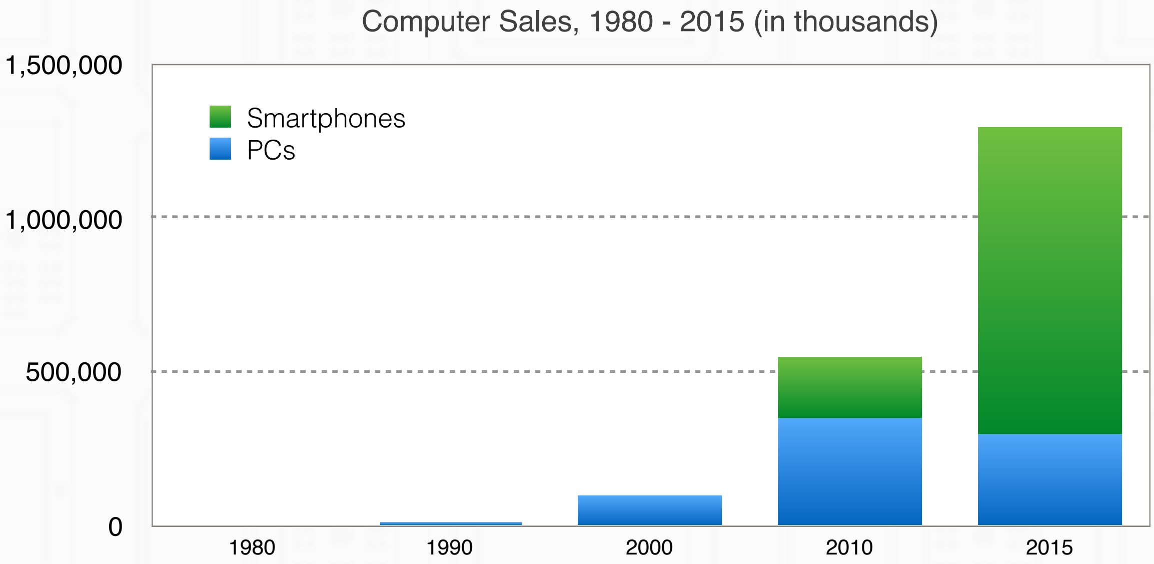
The trend towards smaller and cheaper electronics has been going on for a long time, and in the 1970s it finally produced computers that could be affordable not just for large institutions like the military or General Motors, but for individual consumers.

The Apple 1, 1976 (photo credit: Ed Uthman)

The Apple 1 (shown above), a device handmade by Steve Wozniak, one of the founders of Apple Computer, was arguably the first personal computer, aimed squarely at individuals for home computer purposes. An enormous hit in 1976, it helped to create the consumer computing market that later overlapped with business computing in the form of IBM's famed PC.

[5] http://www.theguardian.com/technology/2012/jul/18/ofcom-report-phone-calls-decline

Though there has been recent speculation and angst about the death of the PC – with sales of desktops and laptops appearing to have peaked in 2010 – it is clear when smartphones are added to the figures (see computer sales graph) that the personal computer has simply gotten much smaller and more mobile.



Big Data Use in Business

Not surprisingly, big data, like many other information technology advances, has proven most productive initially in the realm of business and consumer technology in particular. We are in the middle of a revolution and this is apparent almost monthly, with the appearance of a new app or service that is mining and utilizing our data in new ways. Below are just a few examples of how big data is being used in business and commerce.

Target Knows if You Are Pregnant

As reported in the New York Times[6], statistician Andrew Pole was hired by the retail household goods firm Target in 2002 with a goal of using customers' shopping data to more closely pinpoint what future items they might be interested in purchasing. And an interesting thing occurred:

> As Pole's computers crawled through the data, he was able to identify about 25 products that, when analyzed together, allowed him to assign each shopper a "pregnancy prediction" score. More important, he could also estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy.

Target was able to use this and similar analysis to determine other things about their customers, such as who would respond better to emailed coupons versus paper coupons. They also wrestled with privacy concerns that came up, not surprisingly, when they tried to imagine how customers would respond to Target knowing about their pregnancy when they had not explicitly told Target (or, possibly, anyone).

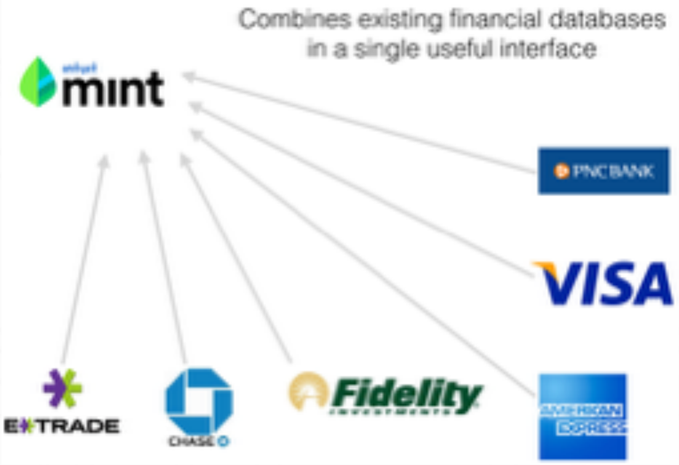
The end result is that Target realized that its databases of client information, detailing purchases, timing of shopping activities, response to emails and mailed coupons, and many other pieces of information, was able to drive much more effective interactions with their customers — leading to improved outcomes (i.e. sales). In effect, they identified previously unknown relationships — between purchases and pregnancy — by analyzing the large datasets they were constantly collecting.

[6] http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

Mint: Unified View of Financial Datasets

Mint.com, launched in 2006, is a financial "aggregation" website: an individual user can provide Mint with login information for his or her financial institutions, and Mint then takes all the information from those institutions and presents them in a unified interface via an interface based in a web browser or a mobile app.

For most people, this will be the very first time they have ever examined their personal finances in a "big picture" way, and it can be a very powerful tool for budgeting, saving, directing investments, and tracking spending.


Combines existing financial databases in a single useful interface
mint
PNC BANK
VISA
E*TRADE
CHASE
Fidelity INVESTMENTS
AMERICAN EXPRESS

Uber Uses Big Data to Improve Transport

The Uber company, based in California, was founded in 2009, two years after the introduction of the iPhone. The founders of the company realized that the widespread use of the new smartphones was generating an enormous stream of data related to the locations of the phones' users. And they further realized that some of those users had a need that other smartphone users could fulfill: transportation.

In effect the entire taxi industry was built to connect potential drivers with potential riders, but born long before smartphones and the internet it relied upon outdated, labor-intensive, and inefficient methods for this.

The three options were:

1. taxis drive around looking for riders
2. riders go to centralized cab locations
3. riders call central taxi dispatchers to arrange for taxis

The problem with 1 and 2 are that they are very inefficient: taxis spent an enormous amount of time vacant, and driving endlessly. And riders had to go out of their way to central taxi locations, even if that location was far from their desired destination.

The problem with 3 is that it was very expensive to employ ten people, for example, to sit around waiting for riders to call.

Uber realized that with the iPhone and other smartphones, an enormous database of potential riders and potential drivers is being constantly updated — and it is much cheaper and more efficient for software to automatically connect a rider with the closest taxi, and a taxi with the closest rider.

Quite correctly, Uber is seen as a challenge and a threat to the livelihood of taxi dispatchers — while in most places being embraced by drivers. No matter what opinion you have of the service, however, it has to be acknowledged that Uber is a much more efficient system than the taxi system it is trying to replace — and this efficiency is based on the utilization of big data.

It's also just as important to remember that there was a short period, between the spread of smartphones and the invention of Uber, where we theoretically had a database of rider locations, and a database of driver locations — but no one was yet connecting those databases. Uber did that, and has dramatically improved local transportation by doing so.

IBM and NIH Working Together with Big Data

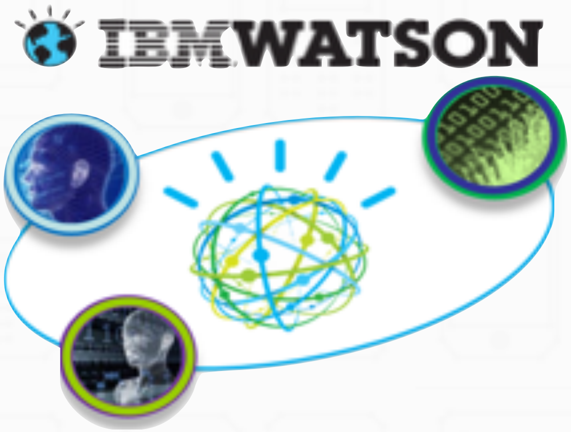
One of the most promising developments in big data has been the advent of IBM's Watson technology, which seeks to allow researchers to much more easily find the answers to questions when those answers require multiple large and complex datasets. Part of Watson's appeal is that it allows researchers to ask questions in "natural language" rather than having to learn a computer query language syntax.

Watson has already been applied to many tasks, and was initially tested in 2008 against human contestants on the TV game show Jeopardy in answering a very wide variety of questions requiring access to very large databases. Watson beat both human champion players.

Since that time, IBM has been seeking to apply this technology within a wide range of medical applications. For example, oncologists at the Maine Center for Cancer Medicine began using Watson technology in 2013 in order to recommend cancer treatments based on large databases of up-to-date information:

> Thanks to its computing power Watson can sift through 1.5 million patient records and histories to provide treatment options in a matter of seconds based on previous treatment outcomes and patient histories. It has been fed with more than 600,000 pieces of medical evidence, 2 million pages of text from 42 medical journals and clinical trials in the area of oncology research, IBM said.[7]

Seeking to capitalize on this groundbreaking approach to making large datasets more accessible, the U.S. National Institutes of Health — the preeminent U.S. government medical research organization, which oversees an annual $31 billion budget — is now working with IBM to connect a very wide variety of clinical and research datasets to the Watson system. This enables dramatically easier and faster querying of the data.[8]



[7] http://in.reuters.com/article/2013/02/08/ibm-watson-cancer-idINDEE9170G120130208

[8] https://ibmecm.cloudant.com/wcm70x/_design/main/_show/detail/ECC-IMC14949USEN?instructions=false&lnk=ushpv18ce2

Conclusion

Many recent technology advances in storage, computing, and networking — along with progress in wearable computing — have been brought together to enable the compilation, combination, and analysis of very large datasets in ways that were simply not possible before.

This has been widely accepted within the business community, with companies as different as Target (a retail chain) and Uber (a transport coordination “app”), with dramatic results (particularly in the case of the controversial Uber) that empower companies and consumers — and that also raise a variety of questions regarding security and privacy. Others are only now beginning to grasp the potential for benefit to consumers from big data tools, and only now beginning to use the tools on offer from large technology companies like IBM, Apple, and others.

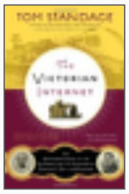
For many fields of industry, this growth will continue to enable new opportunities to use technology to grow and make smarter decisions, whether in the already booming world of retail or just beginning to make a difference in the healthcare industry.

Resources and Further Reading

Big data timeline
http://www.winshuttle.com/big-data-timeline/

Big Data: A Revolution That Will Transform How We Live, Work, and Think
by Viktor Mayer-Schönberger and Kenneth Cukier
http://www.amazon.com/gp/product/0544002695

The Innovator's Dilemma
by Clayton Christiansen
http://www.amazon.com/Innovators-Dilemma-Technologies-Management-Innovation/dp/142219602X/



The Victorian Internet: The Remarkable Story of the Telegraph and the Nineteenth Century's On-line Pioneers
by Tom Standage
http://www.amazon.com/gp/product/162040592X

Big data: The next frontier for innovation, competition, and productivity
by James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers
http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation/