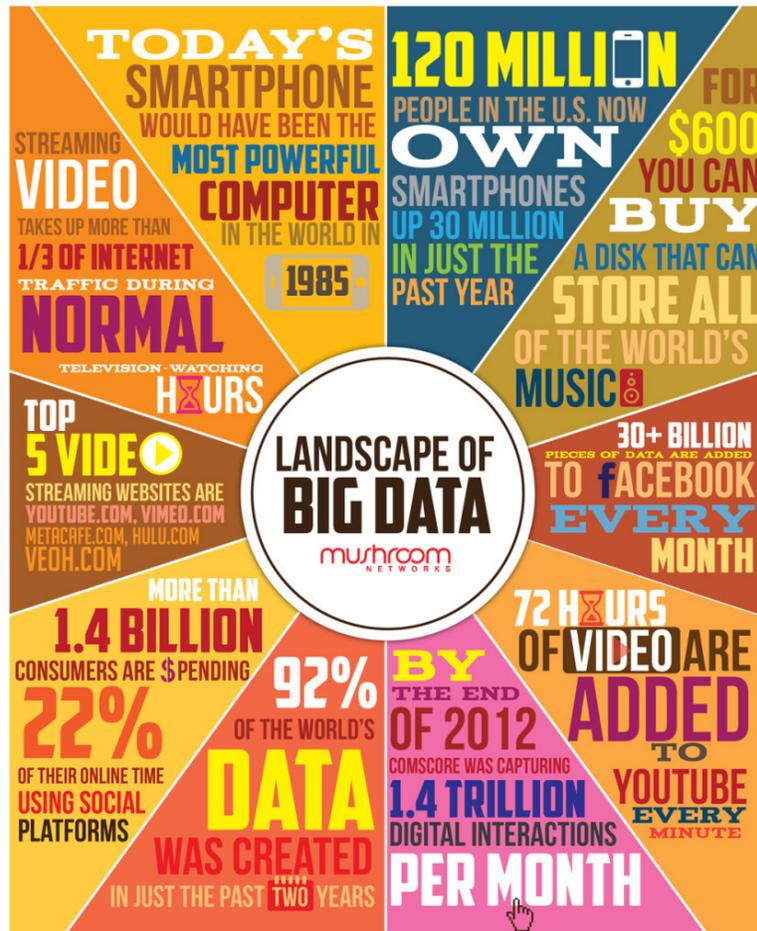


The big deal with Big Data

Author : Cat

Date : January 27, 2014



What is Big Data?

In the beginning, there was simple computing: Think of the most basic computer you have, your calculator. It is great at computing numbers, and displaying the answers.

Then there was number comparison: Comparing a series of numbers requires a spreadsheet, which takes more data and uses graphs and charts for displaying the results. For complex and automated solutions, you would use a computer.

Not good enough! We need analysis and display: Storing more structured information and retrieving it in an ordered way calls for a database and the software to display the results of your queries to the database. For a large scale company that maintains customer data, say a credit card company, they would have need for a dedicated set of servers to store the information, and sophisticated software to pull the right content up when requested.

Which leads to BIG DATA: Now increase the amount of data you have by orders of magnitude. Suddenly conventional servers, software, and processes are no longer effective at storing, analyzing or presenting this data. It is so large in quantity and complexity that we refer to it as BIG DATA.

In 2010 there was 1 ZB of digital data in the world. In 2014, we generated and recorded 72 ZB. A ZB (zettabyte) is 10^{21} bytes: That's a 1 with 21 zeros, or just a very large number. The bottom line is that we produced 72 times more digital data **just last year** than was produced in our human history up the year 2010!

Hey, how did you know that?

Let us imagine that each of us has an electronic file. Ok, not imagination, we do all have a file or multiple files around the world. I am going to call it your BIG FILE of BIG DATA. Some of it is restricted-access and some of it is public.

Somewhere near Bluffdale, Utah, the US government has built the **Intelligence Community Comprehensive National Cybersecurity Initiative Data Center**, aka the Utah Data Center. Its purpose? To store loads and loads of Big Data. It is run by the NSA but regardless of your

nationality, if you have a profile on a US owned application (ex Facebook, Twitter) then you already have a Big File at this location.

The government has always housed a large amount of our data, and not the secret kind. They have all relevant dates (birth, marriage, death), they have every official document we own (passport, license, health card) plus the records of all the times we have moved, where we have worked, and what health treatments we have used. Of course they cannot legally use this outside of government purposes.

What has happened in recent years is the rising popularity of electronic services and cloud computing. Every time you use an electronic device, there is a record of it. And not just social network updates. Think GPS devices in your car, phone calls, loyalty programs and interac payments. Some of this stuff is private but other items are public. They may be kept in the cloud where other companies can also use them. Add to this spending habits (Amazon), travel destinations (trip advisor) and check-ins (foursqaure) and suddenly you have painted a very detailed picture of that person's interests. The data also now comes in multiple forms (text, images, video) and at faster rates than ever before.

Now, the data itself is not the only commodity here. The analysis is more important. Manipulating the data properly allows companies to predict consumer trends and allows advertisers to better target their ads. Social networks, for instance, use predictive algorithms to suggest new members of your network. This means your Big File is half raw data and half analysis which has been inferred from that raw data.

So the end result is that you see curated content. You get a tailored experience that meets your needs and wants. But is it at a price?

Privacy - a trade-off paradox

'Privacy in age of Big Data' has become a hot topic. Mostly because the technology is evolving so quickly that we have not had time or means to put the right privacy protection or legislation in place. If you post something publicly, do companies have the right to take it, keep it and analyze it? How would you answer that question if I asked what you thought about someone riffling through your garbage after it had been picked up by the garbage truck? It feels like an invasion because someone is using your 'stuff' in a way you didn't expect BUT once you release something to the public, what rights do you still have?

Technology comes with all sorts of terms and conditions on their use. Often contained in these terms is the ability to use your data. Here we find the paradox - the software we use that should be FREE has no monetary price tag, but there seems to always be a trade-off as a price.

- The government does not need access to people's personal email or phone calls; but then they cannot guarantee safety for a national threat
- A free app from WebMD helped pregnant women track their pregnancies; but it kept all the data entered (anonymously but with relevant stats) to give to other companies
- The Aerogold VISA allows you to collect aeroplane miles on your purchases; but then gives that purchase info back to Aeroplan for their own use
- Netflix asks you to rate movies so it can make better suggestions; but now it has a list of what you watch mapped against your name. Add that to your Big File!

Privacy is not a black and white area. Different people can have different standards and expectations for where their information is taken and used. It is up to you to decide if you are comfortable with the level of use. **Ask yourself: Does the value of the application, device or service outweigh the value of my data that might be given away?**

Here's one that is beyond my level of comfort: Jack Gallant, a neuroscientist at Berkley, is using "big data" stats to predict what the mind is thinking based on magnetic resonance imaging measurements. You wear a device on your neck that measures how your brain reacts to different situations. When confronted with a decision, the device has determined how you might react and then it can send a signal to either steer you towards your usual decision or away from it. Imagine the implications of actually controlling how people think. Too science fiction for my liking.

DON'T PANIC!

Sometimes I feel overwhelmed with all the potential cases of 'misuse' of my data. Yet, I am also excited about the possibilities with Big Data! Using the data to understand the customer also makes for a better product. When I shop online, I do not want to sift through an entire catalogue when I could be presented only with the items I would actually buy. And maybe all of this data will help governments to make better choices that serve the good of the people.

Too far fetched? Possibly. The US government has at least launched a group to help define some

of the rules and processes. According to John Podesta, a member of the committee, it is a "comprehensive review of the way that big data will affect the way we live and work; the relationship between government and citizens; and how public and private sectors can spur innovation and maximize the opportunities and free flow of this information while minimizing the risks to privacy." They will be consulting members of industry, academic organizations, privacy groups and civil liberty groups. The results of this committee are expected end of April, 2014 and will be published after that.

TL;DR (Too long; Didn't Read)

Big Data is the storing, analysis and presentation of enormous quantities of data that are created in every moment and by every electronic device. A Big File exists that contains all *your* correlated data. Typically you are trading your information for any free service you use. Governments are working on legislation to protect privacy, but it does not exist yet.