# The Big *Boom* of *Big Data*

Ricardo Fusco

Universidade de Évora, Portugal
`ricardo.fusco2@gmail.com`

**Abstract.** Nowadays we find ourselves in the golden age of *Big Data* with this concept being a rising trend. Given the exponential rise in the amount of data being generated daily by all kinds of companies, it was inevitable to stumble upon a whole new array of markets with the capture and analysis of all types of data in order to provide a full picture of most of the business related aspects and adapt to the ever changing needs and behavior of costumers. Along with all the new technologies available to deal with large volumes of data came *Data Warehousing* and *Business Intelligence* [CCS12] completely revolutionizing the IT industry.

## 1   Introduction

A few years ago it was almost unthinkable to try dealing with big amounts of data due to the limitations regarding storage and CPU technologies reaching the point where scalability became a serious problem. Then came along Moore's law tipping the scales for the IT industry resulting in the development of more efficient, effective, robust and faster storage and CPU technologies. This law basically states that the processing power/capacity and processing speeds, more specifically the number of transistors in the processor, double every two years [moob] [Sha14] [Rus11].
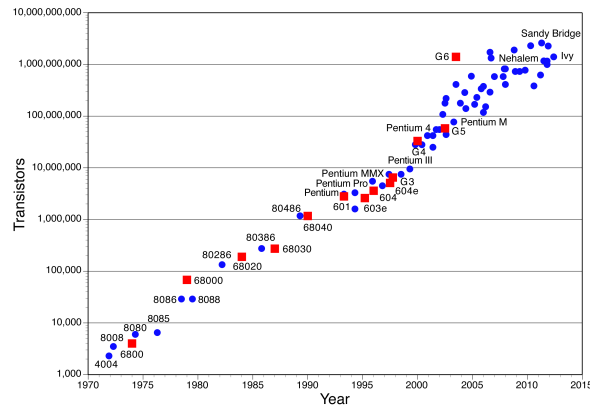


**Fig. 1.** Moore's Law [mooa]

This all lead to today where companies seek new and inexpensive ways to acquire and analyze data in order to gather facts, new information, basically data about the data per se. What is being revolutionary in the *Big Data* era is not so much the enormous amounts of data being generated but what one can do with it. Furthermore, with the amount of technologies and tools now available and still being created, it is now possible to more efficiently and effectively address the needs of the costumers and improve the user experience. [Sha14]

## 2  Big Data

The *Big Data* phenomenon is mostly due to the continuous advances being made in technology allowing for a deeper understanding of everything and of the best courses of action to take regarding the discovered facts. The pace at which the data is being generated keeps growing at an astounding rate every single day. And it is not merely just about the increasing quantities of data but also about the new types of data being created, like, for instance, data gathered by all kinds of sensors.  [KF13] [Rus11]
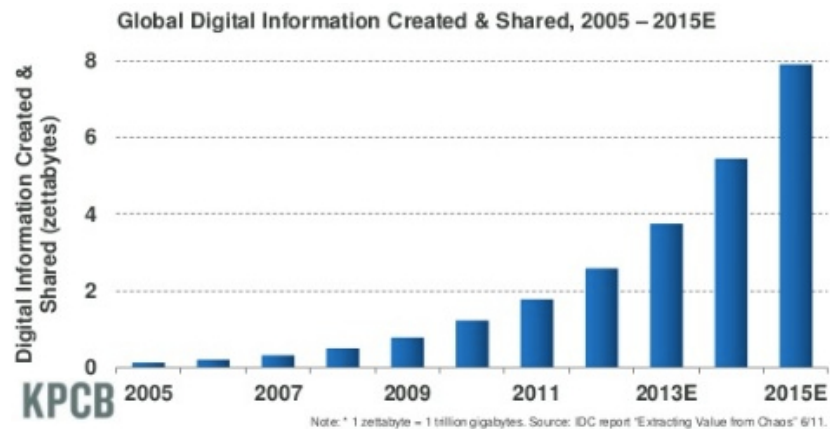


**Fig. 2.** Digital content generated and shared per year (zettabytes) [Bye]

Data availability is less and less an issue, most of which is random bits and pieces of meaningless data which must be properly processed in order to make some sense out of it. All of this adds up to the *Big Data* revolution. Things like business intelligence, artificial intelligence techniques like machine learning, pattern recognition can greatly benefit from this accelerating furthermore the evolution in computational techniques.

But it is not all a bed of roses. As *Rumpelstiltskin*, the dark one, is used to saying - "All magic comes with a price." - the same applies to *Big Data*, it also has

its downsides and negative aspects. When working with massive amounts of data, doing thorough analysis of these data, statistics and other kinds of data analysis the risk of reaching wrong or meaningless conclusions is quite considerable, like looking for a small needle in a colossal haystack of data, misleading and leading one to the inevitable fact here that there are "many bits of straw that look like needles". [BC]

Another crucial negative point regarding *Big Data* is the ethical consequences associated with the privacy issues, studies on some sets of data can harm and/or breach one's privacy which can sometimes lead to another problem which is the uncertainty related with the amount of damage one might or might inflict upon someone while using data sets about humans, not only that but there are also "Numerous companies that collect and sell, consumer profiles that are not clearly protected under current legal frameworks,..." [BC] resulting sometimes in illegal, or less legal, brokerage of data regarding subjects. Breaches of data can be equally damaging to one's privacy, for example, breaches to someone's medical records, social records, etc. The progress and developments made in the future regarding the privacy/anonymity issue remain to be seen. [BC] [Tay14] [Loh12]

## 3    Fields/areas where *Big Data's* potential is being explored

Taking as an example social networks like *Facebook*, *Twitter*, *Flickr*, *Instagram*, amongst others, we can closely observe how the parsing, processing and measurement of the billions of billions of data generated by these entities can provide information about its users like behavior patterns, potential marketing material and information on how people are connected to each other in an extremely detailed way. [Loh12]



**Fig. 3.** Facebook's global relationships visualized in a global map [J.10]

In the economic sector it is more and more frequent to see decisions relying on the results provided by the data analysis more so than on experience or induction. Big-box stores and retailers can now thoroughly analyze their prices, successful and unsuccessful sales, and other kinds of data about their costumers. [Loh12]

Another clear example of how this *Big Data* trend has been evolving is the oncology sector, with the lowering cost of gene sequencing and the development of the industry making gene sequencing easier, genomic data seems to be getting more and more bountiful and turning gradually more crucial to the development and application of cancer treatments. Clinical organizations and biotechnology companies generate large quantities of genomic data through their research, and quite a lot of organizations have developed initiatives for generating and sharing genomic data from cancer patients and cell lines with the public, which can then be processed and parsed using a *data warehouse* given the need here for data management. [Tre14] [Loh12]
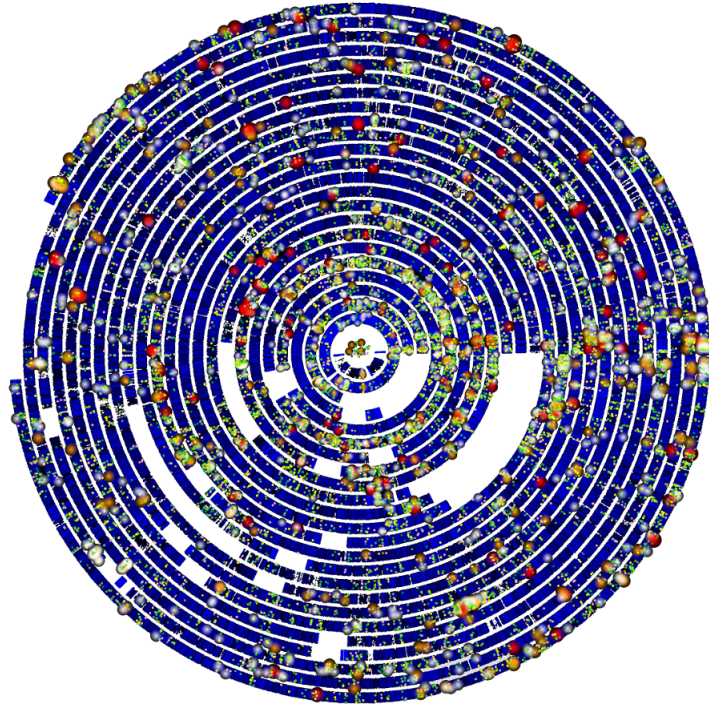


**Fig. 4.** Genomic data ("Everybody's uniqueness"). Image by Beat Wolf, Pierre Kuonen and Thomas Dandekar  [Tre14]

## 4  *Big Data* Technologies

Nowadays most companies try to turn to the more cheap and efficient technologies in order to treat and make something out of the gathered data, more specifically NoSQL solutions. Why? Well because NoSQL is one of the base groups of *Big Data* technologies, a new and more adequate solution for solving problems that the regular relational database management systems (SQL) seem to have regarding these large amounts of data, like scalability, NoSQL technologies were specially made to be easily scalable thus being cheaper and easier to work with unlike relational databases which are much more complex and costly in terms of servers, licences, extended periods of time for updating and providing new services, etc. Relational databases also have problems dealing with some of the most common data types in these cases like, for example, "unstructured data, blobs of text, sensor data ...". [Hir14] [KF13] [Dix14]

One of the several examples of a successful investor in NoSQL technologies is Netflix going from just another boring DVD rental company to one that streams daunting amounts of popular content to millions of viewers generating trillions of transactions per day. [Hir14]

If the company seeking for *Big Data* technologies happens to have quite deep pockets then it can, for instance, spend its money on an expensive *Big Data* solution from IBM like some of the largest companies did, like the well known retailer Barnes & Noble, like Barclays, Mc Donalds UK, within Portugal, Banco Espirito Santo (now named Novo Banco), amongst others. [IBMa] [Hir14] [Dix14] [Oli]

Besides NoSQL, Hadoop seems to be a quite valid tool for *Big Data*. This technology is a software framework created in order to "help users handle compute-intensive processes on server clusters with large data volumes. With the help of Hadoop, applications can distribute complex computing tasks across thousands of nodes and process data volumes in the petabyte range ..." [Kon] proving to be an essential tool along with NoSQL for accessing and interacting with substantial amounts of data.  [IBMb] [KF13] [Oli] [Mit14] [Lo] [Rus11]

## 5  Conclusion

*Big Data* does not comprehend only the data volume, its size whilst in storage, it goes beyond that encompassing also two more properties, speed at which data is processed and data types' diversity.  [Rus11]

There are lots of sectors/areas which benefit greatly from the data analysis, using data warehouses and applying Business Intelligence techniques to further improve one's business. There are also the downsides to this trend, which will be interesting to see how it plays out. The *Big Data* topic however remains a quite "hot" one these days and seems to be an ever growing phenomenon, it brought, and surely is yet to bring, a lot of change and improvement to most of

the different sectors/areas. It will be interesting to see how this "boom" of *Big Data* unfolds in the future.

## References

[BC]     Danah Boyd and Kate Crawford. CRITICAL QUESTIONS FOR BIG DATA. (5):20.

[Bye]    Kleiner Perkins Caufield Byers. Global Digital Information Created & Shared (Image). http://heidicohen.com/content-marketing-7-steps-to-a-successful-repurposing-strategy/.

[CCS12]  Hsinchun Chen, Roger H. L. Chiang, and Veda C. Storey. Business Intelligence And Analytics: From Big Data To Big Impact. 36(Special Issue: Business Intelligence Research):25, 2012.

[Dix14]  John Dix. What's better for your big data application, SQL or NoSQL? http://www.networkworld.com/article/2226514/tech-debates/what-s-better-for-your-big-data-application--sql-or-nosql-.html, 2014.

[Hir14]  Scott Hirleman. Big Data or Big Hype or Why NoSQL Matters, 2014.

[IBMa]   IBM. Business Case Studies. http://www-03.ibm.com/software/businesscasestudies.

[IBMb]   IBM. Hadoop system. http://www-03.ibm.com/software/products/en/category/SWP12.

[J.10]   Lee J. Facebook's global relationships visualised in a natty map. www.wirefresh.com/facebooks-global-relationships-visualised-in-a-natty-map/, 2010.

[KF13]   Jeff Kelly and David Floyer. The Industrial Internet and Big Data Analytics: Opportunities and Challenges. http://wikibon.org/wiki/v/The_Industrial_Internet_and_Big_Data_Analytics:_Opportunities_and_Challenges, 2013.

[Kon]    Ralf Konrad. Emerging technologies: Hadoop and NoSQL Accessing and utilizing large amounts of data. http://www.t-systems.com/solutions/emerging-technologies-hadoop-and-nosql-accessing-and-utilizing-large-amounts-of-data-/1146808.

[Lo]     Frank Lo. Big Data Technology. https://datajobs.com/what-is-hadoop-and-nosql.

[Loh12]  Steve Lohr. The Age of Big Data. *The New York Times*, pages 1–5, 2012.

[Mit14]  Robert L. Mitchell. 8 big trends in big data analytics. http://www.computerworld.com/article/2690856/8-big-trends-in-big-data-analytics.html, 2014.

[mooa]   Moore's Law graph. http://www.overclock.net/image/id/10100141.

[moob]   Moore's Law. http://www.mooreslaw.org/.

[Oli]    Andrew C. Oliver. The dirty truth about big data and NoSQL.

[Rus11]  Philip Russom. Big Data Analytics. *October*, 19:40, 2011.

[Sha14]  Jonathan Shaw. Why "Big Data" Is a Big Deal. http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal, 2014.

[Tay14]  Taylor Armerding. The 5 worst Big Data privacy risks (and how to guard against them). http://www.csoonline.com/article/2855641/big-data-security/the-5-worst-big-data-privacy-risks-and-how-to-guard-against-them.html, 2014.

[Tre14]  Virginia Tressider. When it comes to big data, the eyes really have it. http://csironewsblog.com/2014/09/29/when-it-comes-to-big-data-the-eyes-really-have-it/, 2014.