The Big *Boom* of *Big Data*

Tópicos Avançados de Bases de Dados 2014/2015

Ricardo Fusco

Escola de Ciências e Tecnologia
Universidade de Évora

April 27, 2015

UNIVERSIDADE DE ÉVORA

Summary

UNIVERSIDADE DE ÉVORA

What is it? How did this phenomenon start?

- A few years ago it was unthinkable to deal with large amounts of data.
- There were limitations regarding the hardware.
- The evolution regarding hardware and technologies propelled the development of several industries to a new level.

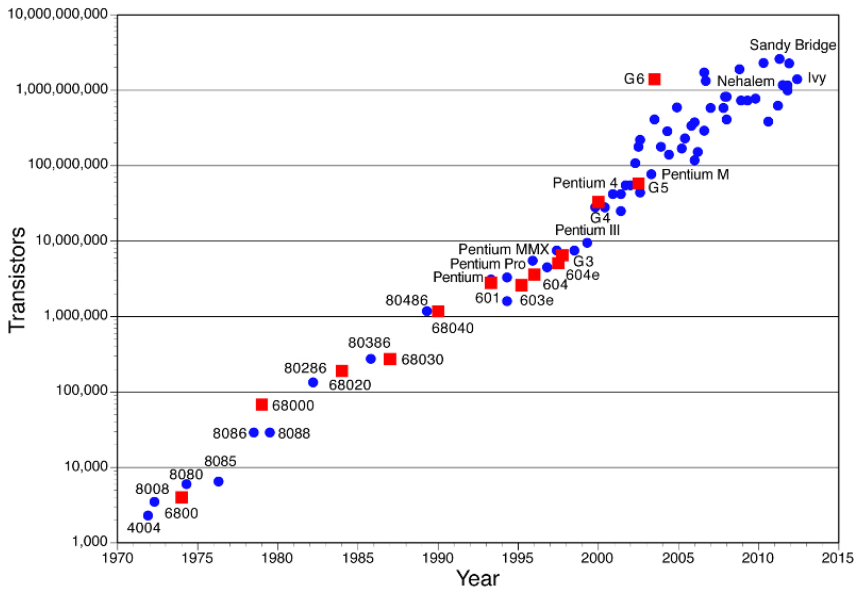

Figure 1: Moore's Law

UNIVERSIDADE DE ÉVORA

What is it? How did this phenomenon start?

- Continuous advances in technology allow a deeper analysis and understanding of the data.
- One of the interesting parts about *Big Data* is, not only the massive quantities of data, but also the growing number of new types of data.
- The pace at which data is being generated keeps growing at an astounding rate.

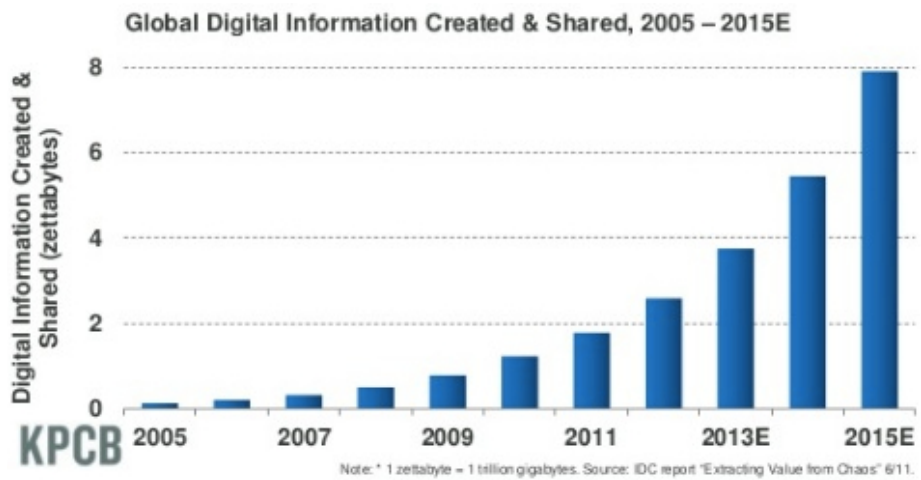

Figure 2: Digital content generated and shared per year (zettabytes)

UNIVERSIDADE DE ÉVORA

Positive aspects

The capture and analysis of data adds a great value to any industry and/or company who decides to harvest its benefits. Every aspect, or at least most of them, will eventually end up being affected by *big data*

- New industries and areas of expertise are created.
- A great new way to understand and target customers.
- Understand and improve the business process and model.
- We can all, individually, benefit from the insights the data generated by our devices and sensors provide.
- Improvements in science, research and also regarding machines' and devices' performance.
- Improvements, like for instance, better decision making in the economy sector.
- Better public health and healthcare services
- . . .

UNIVERSIDADE DE ÉVORA

Negative aspects

As there are positive aspects to it, there are also negative aspects and downsides when working with huge data sets that may sometimes generate setbacks in the struggle to take meaning out of this data.

- The risk of reaching wrong/meaningless conclusions is somewhat high.
- There are also several ethical problems mostly related to the privacy issues involved.

UNIVERSIDADE DE ÉVORA

Social Networks

- Social networks like Facebook, Twitter, Flickr, etc, generate billions of billions of data.
- Lots of information about people, behaviours, etc, can be retrieved for several ends.

Figure 3: Facebook's global relationships visualized in a global map

UNIVERSIDADE DE ÉVORA

Economy

- It is becoming more and more common to use big data algorithms in the financial trading area.
- It is also very beneficial to optimize ones' business.
- Allows for a deeper glimpse of what to improve financially, say in a company for instance.

Figure 4: Global economy

UNIVERSIDADE DE ÉVORA

Bioengineering: Oncology

- Due to the enormous amounts of genomic data being increasingly generated it is essencial to treat that data.
- There is a clear need for data management in the oncology sector.
- Allows for a deeper glimpse of what to improve financially, say in a company for instance.
- It greatly improves the development and application of cancer treatments

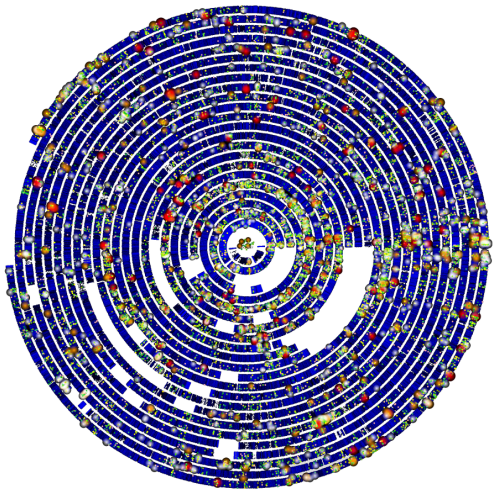

Figure 5: Genomic data ("Everybody's uniqueness")

UNIVERSIDADE DE ÉVORA

NoSQL and SQL

- NoSQL is one of the cornerstone groups of big data technologies.
- One of its best features is scalability and it allows for a less complex and less expensive way to store and analyze data.
- Netflix is an example of a successful user of NoSQL technologies.
- There are also several SQL alternatives for big data, for instance the ones provided by IBM.
- These SQL alternatives are more complex and expensive for big data.

UNIVERSIDADE DE ÉVORA

Hadoop

Hadoop is also an extremely useful technology for big data purposes.

- It helps compute intensive processes on clusters.
- Complex tasks can be efficiently distributed across thousands of nodes.
- Allows to more easily process petabytes of data.

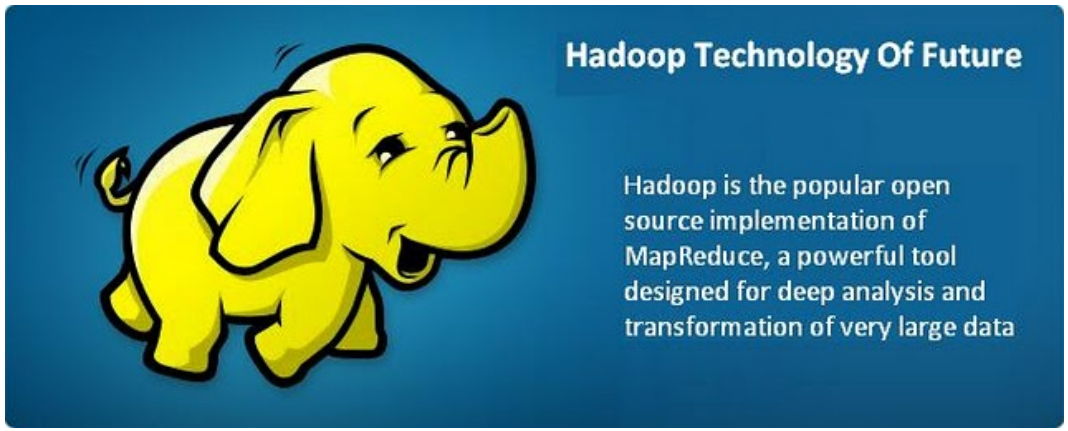

Figure 6: Hadoop

UNIVERSIDADE DE ÉVORA

Conclusions & Future Perspectives

- Big Data came to solve lots of problems and has the potencial to solve much more.
- The positive aspects about this technology seem to greatly outweight the negative ones.
- Data warehousing and business intelligence are becoming more and more important in this big data revolution.
- Remains to be seen the developments regarding the ethical and privacy issues regarding big data.
- It will be interesting to observe how this whole big data *boom* unfolds.

UNIVERSIDADE DE ÉVORA

Conclusions & Future Perspectives

The End

UNIVERSIDADE DE ÉVORA