The business of information

interlogica.it

interlogica

> Information is the oil of the 21st century, and Analytics is the combustion engine
>
> — Peter Sondergaard, Senior Vice President, Gartner

Summary

BIG DATA
A short introduction

Big Data and Analytics are two of the biggest revolutions happened in the last few years in the digital world. It is a growing trend which will have an even bigger impact on our lives but it will also revolutionalise the way in which we do business.

Some of our daily actions generate a huge number of data: interconnected smartphones, clicks on websites, smart meters, IoT devices and vocal requests to Google, Siri or Alexa. These are massive volumes of miscellaneous data from different sources and with different formats which can all be analysed in real time: all this is Big Data.

Big Data is quite a new concept that still now struggles in getting a proper and standard definition. The first description of Big Data is linked to the size of data, whose starting point is measured in terabytes (1024 gigabyte), petabytes (1024 terabytes) and even more, using zettabyte (billions of terabytes). To give the concept some tangible sense, experts have divided it into 3 main relevant segments, calling them the 3V of growth for Big Data:

Variety - all the information archived in Big Data are tremendously different and each one of them comes from a very specific source.

Velocity - extreme speed is needed to archive, save and catalogue such extended amount of information.

Volume - technological solutions are needed to allow to have all the massive dimensions (from a bytes perspective) of Big data managed.

These fundamental features have now divided into some further 10 segments – the 10V of Big Data – to describe the whole concept in full details.

When we talk about Big Data we need to understand the changes that they involve: from the data gathering process, to the technologies supporting the lifecycle of data and the development of new competences for valorizing the single datum. Managing and analyzing Big Data require using appropriate technologies and having professional experts who can optimize availability and usefulness. Hence why the Data Scientist becomes a key

Market value

1,4 MILLIONS OF €

of large companies introduced Machine Learning skills..

Big enterprises data

44% Use Analytics Real-Time or Near Real-Time.

56% Already have a Data Analyst on the staff.

46% Already have a Data Scientist on the staff.

42% Already have a Data Engineer on the staff.

figure in the process.

The information gathered from the analysis of such data helps to direct the business actions: marketing experts can then create strategic campaigns, health workers can easily identify epidemics and environmentalists can fathom future sustainability and so on.

Seeing all this from a market perspective, it is then possible to obtain a huge competitive advantage thanks to the ability of making prompt and well informed decisions – this works for big organizations but it also works for small and medium companies.

According to a report published by Assintel about ICT market trends in March 2019, a growth of 1.6% is expected to happen in Italy for the whole year. The incremental trend is guided mainly by big and medium enterprises which are investing in technologies known as "Innovation Accelerators", such as Internet Of Things, Artificial Intelligence, Big Data and Analytics, which are evolving at a very fast pace.

Comparing such info with a market which is globally and constantly growing by 1.4 billions Euro, the Italian situation of Big Data is definitely not rosy. The biggest percentage of investments comes from big companies. A third of these companies has introduced competencies on Machine Learning, 44% use Analytics Real-Time or Near Real-Time. 56% is already employing Data Analysts, 46% has Data Scientists and 42% Data Engineers.

Only a very tiny percentage of medium and small organizations has decided to start Big Data Analytics projects. This is quite a depressing fact considering that the economic market is mainly led by those medium and small organizations.

Origins of data

INTERCONNECTED SMARTPHONE

SOCIAL

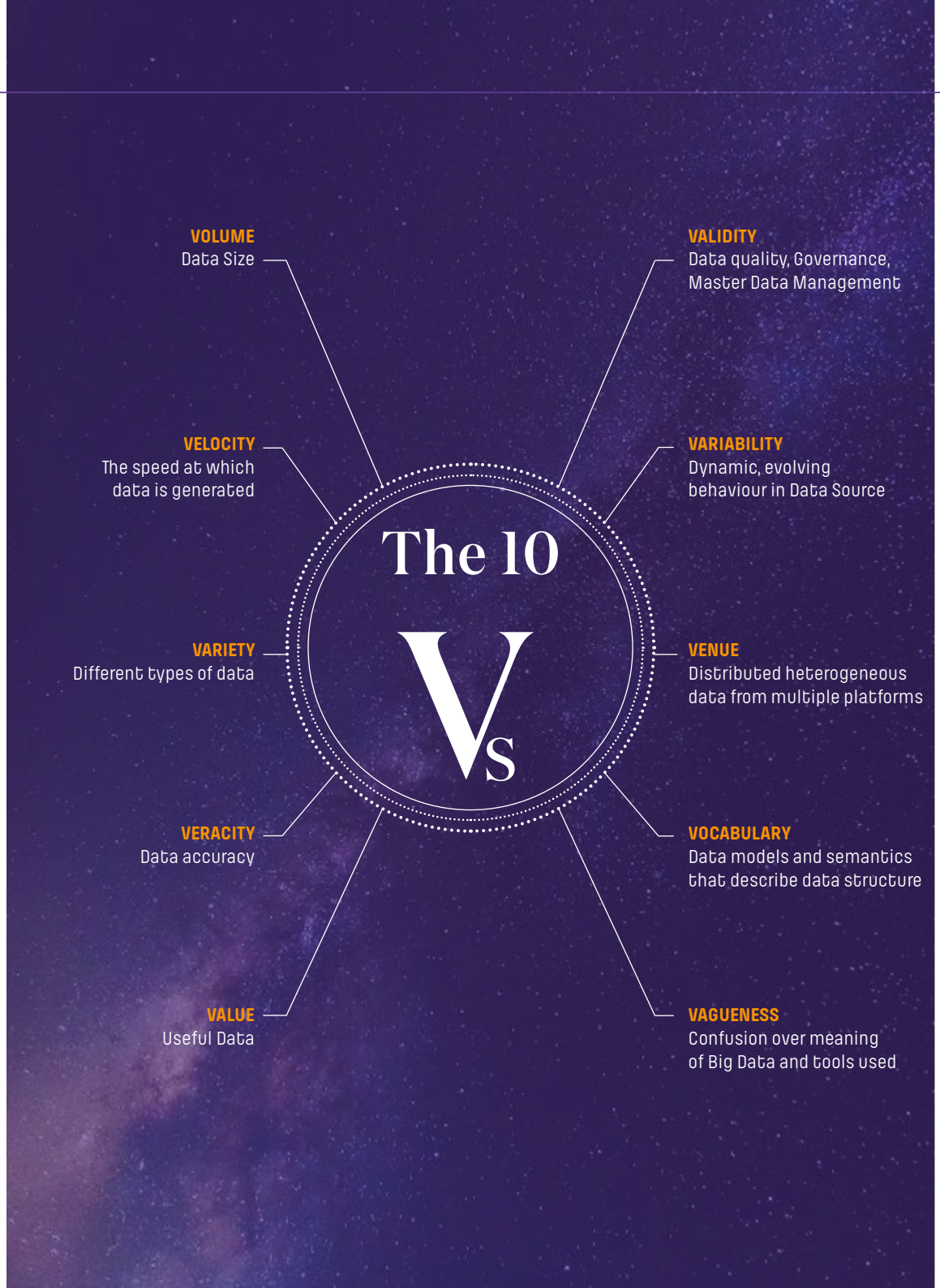
CLICK ON A WEBSITE

SMART METER

IOT DEVICE

GOOGLE RESEARCH

VOCAL RESEARCH ON GOOGLE, SIRI OR ALEXA



Data weight

ZETTABYTE
10^{21}

EXABYTE
10^{18}

PETABYTE
10^{15}

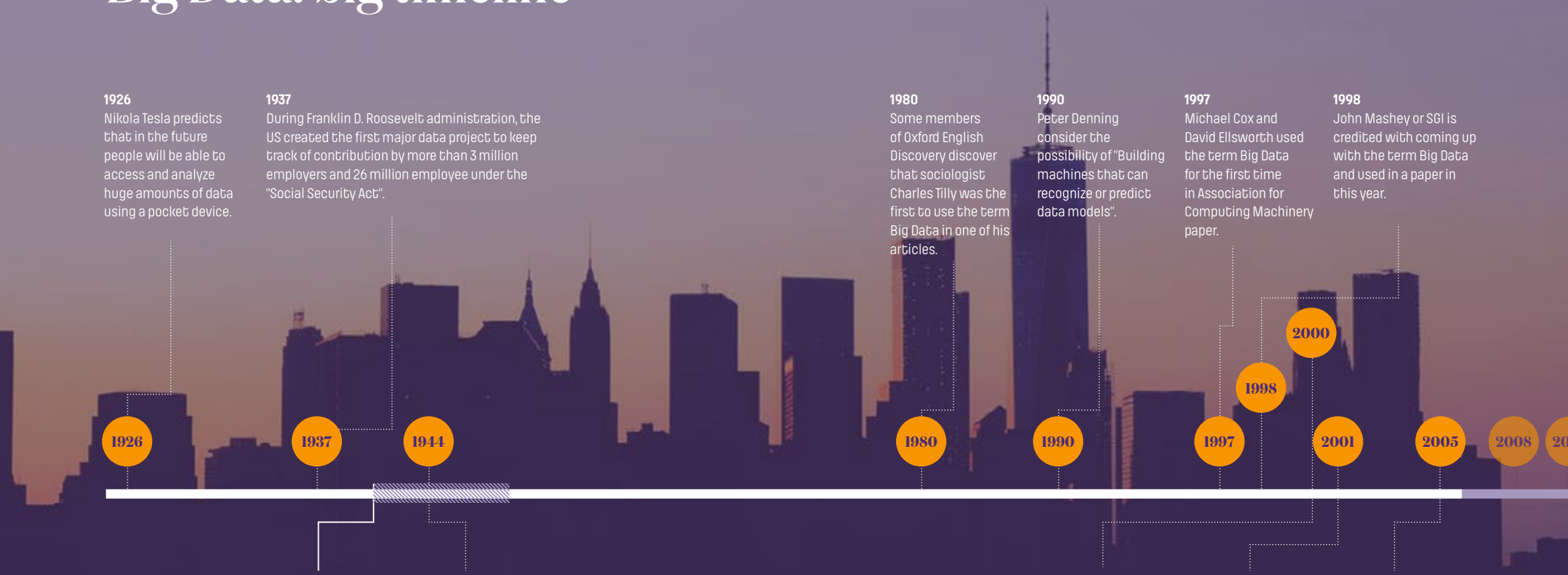
TERABYTE
10^{12}

GIGABYTE
10^{9}

MEGABYTE
10^{6}

CHILOBYTE
10^{3}

Big Data: big timeline



1926
Nikola Tesla predicts that in the future people will be able to access and analyze huge amounts of data using a pocket device.

1937
During Franklin D. Roosevelt administration, the US created the first major data project to keep track of contribution by more than 3 million employers and 26 million employee under the "Social Security Act".

1940's-1950'
The electronic calculator is born to perform high-speed calculations.

1944
Fremont Ryder, librarian of Wesleyan University, hypothesizes that, due to the explosion in the amount of information, the Yale Library in 2040 will have 200 million volumes.

1980
Some members of Oxford English Discovery discover that sociologist Charles Tilly was the first to use the term Big Data in one of his articles.

1990
Peter Denning consider the possibility of "Building machines that can recognize or predict data models".

2000
Francis Diebold referred to Big Data as "the explosion in the quantity (and sometimes, potential quality) of relevant available data".

1997
Michael Cox and David Ellsworth used the term Big Data for the first time in Association for Computing Machinery paper.

2001
Doug Laney (Meta/ Gartner) came up with the 3 'V's (Volume, Velocity, Variety).

1998
John Mashey or SGI is credited with coming up with the term Big Data and used in a paper in this year.

2005
Hadoop was created by Yahoo! built on top of Google's MapReduce. Tim O'Reilly published "What is Web2.0?" Roger Mougalas of O'Reilly Media used the term 'Big Data in its modern context.

2008
Google processed 20 Petabytes of data in single day.

2010
Google's executive chairman, Eric Schmidt, states: "As much data is now being created every two days as was created from the beginning of human civilization to the year 2003".

2014
88% of business executives believe Big Data analytics is a top priority for their business.

2018
Forrester Research concludes, "70% of enterprises expect to implement artificial intelligence (AI) over the next 12 months, up from 40% in 2016 and 51% in 2017". Additionally, it predicts, "50% of enterprises will adopt a cloud-first strategy for big data analytics," and more than 50% of chief data officers (CDO) will now report directly to their CEO.

2020
Data traffic: Big Data grow to an annual rate of 40%..

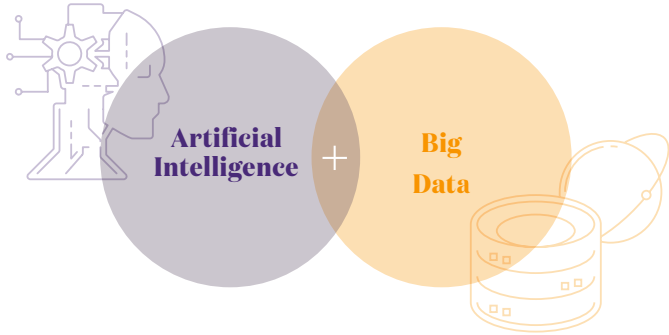
2022
Cisco in the twelfth edition of the study "Mobile Visual Networking Index" claims that the advent of 5G combined with the growing number of devices and connections will lead to a consequent increase in data volume.

In the Italian mobile networks there will be 165 million devices and IoT connections, with an increase of 11.1% compared to 2017, when there were just over 97.6 million.

2021
BCC Research estimates that the market value of the Smart City will reach 775 billion dollars this year, growing to an annual rate of 18%.

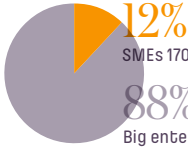
2015
Big Data guru Bernard Marr prognosticates to industry and government leaders that "Big Data is not a fad," that the wind of change is upon us. "We are just at the beginning of a revolution that will touch every business and every life on this planet," he wrote. He, of course, was right. Each year since, some major media outlet has proclaimed that year as the "Year of Big Data," because it was.

2016
About 3 billion people in the world have web access. Their activities and preferences - terabytes of data - are largely gathered to be subsequently explored and used, quantifying and measuring collective behavior to understand and predict the functioning of our society and market trends.



2018 Analytics Market*

1393
MILLIONS OF €



12%
SMEs 170 mln€

88%
Big enterprises 1223 mln€

Big Data adoption trends

> The digital transformation inside companies enabled by Big Data Analytics can be seen in different aspects which involve technology, processes, organisation and company culture.

2019 has definitely seen a massive growth for the Big Data Analytics in the Italian market. Companies are now finally starting to understand its potential for strategical opportunities.

Big Data are fundamental to carry out detailed analysis to understand, help and support decisional processes and competitive advantages in all the production areas. The market evolution brings customers to demand even more smart services which can solve multiple issues at the same time and in a super-fast way.

Companies themselves are starting to suffer from the global competition and they do need to rationalize expenses while following a transformation path which will turn them into data-driven companies. In such way they can offer more effective and fast answers to the constantly growing demands of the business.

Small and Medium Enterprises (SMEs) still own a small 12% of the market since the real growth is made by big companies (with more than 249 employees) and they represent the remaining 88%.

For SMEs, the biggest problem can be identified in the scarce understanding of the subject, highlighted by the typical heterogeneity of Big Data and by a technological immaturity which doesn't allow to benefit of all the different project advantages they offer.
The main issues can be often found at the very roots of a company: there is a lack in the way data are stored and no strategy can therefore be implemented. There is also a real lack in understanding the value of Big Data business and an approach to Data Analysis which is now obsolete and limited.

For big enterprises the biggest issues arise in the lack of competence and of subject matter experts, in the way data are integrated and in predicting the benefits which come from the investment (Source: Data from the Big Data Analytics & Business Intelligence Observatory of the School Management of the Milan Polytechnic).

Whilst analyzing the awareness and the technological maturity of Small and Medium Enterprises, we can assess that 10% of them is still struggling in understanding the advantages Big Data bring and how the approach to Data Analysis is still limited and traditional.

The new challenges are therefore happening on the speed and intelligence of Fast & Smart solutions:

1- Real time analysis of integrated data from different sources to rapidly respond to the business needs.

2 - Need of intelligent learning mechanisms (such as Deep Learning and Machine Learning) which allow the collection of patterns and data which can bring in a competitive advantage.

* Big Data Analytics & Business Intelligence Observatory of the School of Management of the Milan Poytechnic

Curiosities

#1

A man enters a Target shop, just outside Minneapolis and asks to speak with the manager. What was the purpose of the visit? Well, according to some people, he had quite an intense complaint to make because his daughter, who was still at high school, was receiving flyers and coupons from Target about maternity and baby clothing. "Are you trying to make sure she gets pregnant?". The shop manager apologised and he was so mortified that he even phoned the man at home to say how sorry he was another time. Over the phone it was actually the father himself whom apologized to the manager saying: "I spoke with my daughter, I didn't know it but she is pregnant".

The New York Times wrote an article about such fact in 2012.

Andrew Pole, one of the Marketing and Statistics managers at Target, explained that such situation was the result of studying clients' behaviours using Big Data, just by observing the buying habits and crossing them with other demographical information. Pole and his colleagues can predict clients' behaviours and use them.

Going back to the gentleman from Minneapolis, Pole and his team were able to link the possible pregnancy to the shopping made by the daughter as she had started to purchase products which are usually bought by mums to be (such as certain types of creams, calcium, magnesium and zinc-rich products and others).

Such story really does explain how Big Data and Marketing allow to predict consumers' needs but also how they can direct the shopping habits.

#2

During WWII the British Government asked Abraham Wald, an expert of statistics, for help. They wanted to study and understand how to make planes less vulnerable.

Wald started to observe when planes were coming back from the different missions and created a map to study where bullets hit the planes most.

The conclusion of such work was to strengthen the areas which were not damaged as, since Wald concluded, they still made it back to the base, even though they had been hit. The fact that those parts had not been damaged and allowed the plane to return home was an important suggestion to make them safer and more impenetrable to make sure the plane was safe.

Such situation is a "vintage" example of Survival Bias, also known as Survivorship Bias, which brilliantly shows how it is possible to make informed and data-driven decisions.

The link between Big Data and Artificial Intelligence

"The expression "Big Data" is the popularisation of Machine Learning techniques: computers work on massive amounts of data to understand links and tendencies which human beings fail to see. This explains all the big discoveries we have now achieved in the last few years".

Kenneth Cukier – Senior Editor Digital Products , The Economist

As already highlighted before, every single organization can access a big amount of data: information about the business, from the web, social media, IoT sensors and from other sources which are now extremely important for gathering details used to help the business grow. The simple gathering or access to data, without implementing the correct analysis, produces no results and ends up in being just a huge pile of raw data.

To transform all these data into something very useful we need to synchronize technologies. The connection between Big Data and Artificial Intelligence is fundamental to transform the initial data in information which can predict and guide us towards more effective decisions.

The automation of decision has to absolutely go through the merge of technologies to be agile and fast to satisfy the market of today.

The centre of the process is to become a data driven organization, a situation which is often happening in big realities where decision making processes are driven by more refined data.
This allows to have decisions reached in a short time and they are based on the robustness of objective and systematic evaluations.

Up till now, companies have never really given much attention to understand processes and clients' behaviors and have never measured nor used the data they had in a satisfactory way. This is the main reason why small and medium size companies deeply believe Big Data is something quite useless.

How can raw data be transformed into something which gives added value to the business?
Many organizations are now working on how to transform "raw data" in high potential AI fuel. Info gathering, archiving and Big Data elaboration still suggest some significant challenges for all those firms who cannot afford to employ experts who will look after it.

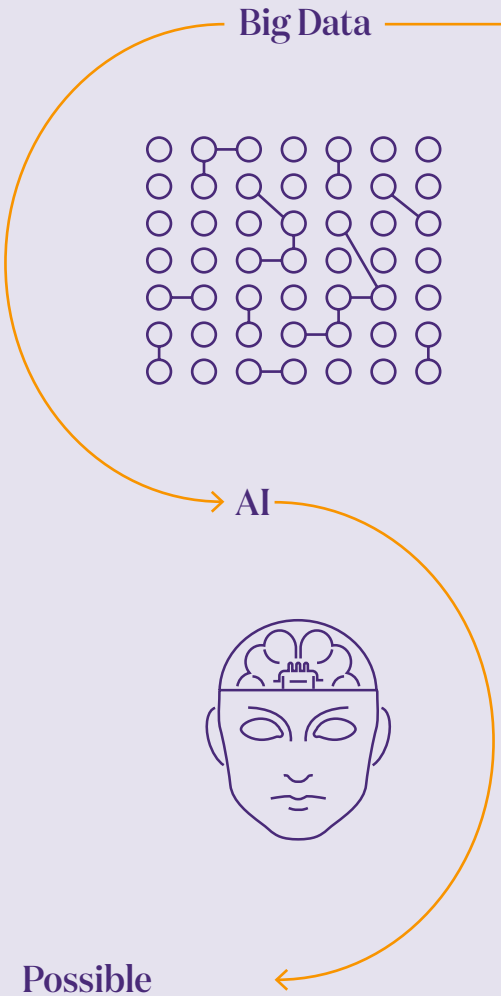
Analytics on Big Data allow to efficiently optimize many different business aspects, for example helping and increasing rapidity in how to analyze and visualize big amounts of data, something which constitutes the first fundamental step towards the appreciation of Big Data.

It is now important to clarify two concepts which are usually associated with one another: Big Data is the raw material whilst Artificial Intelligence (AI) provides the tools to manage and better understand it.
There are different uses which result from the Big Data – Artificial Intelligence interaction: we can analyze customer sentiment and help marketing to drive focused actions; market trends can be identified before they are even born so the company can be driven towards a strategic decision-making process, competition can be assessed, analyzed and studied too and much more.

The totality of big organizations adopt descriptive Analytics but many of these companies are also experimenting an evolution towards Predictive, Prescriptive and sometimes Automated Analytics logics. These are ways of using Big Data to foresee or simulate future scenarios in an automated way. The transformation happens by techniques of Machine Learning and Deep Learning which habilitate types of analysis of Real-Time Analytics and Artificial Intelligence.

Here at Interlogica we create solutions for managing, analyzing and visualizing (structured and unstructured) Big Data from every existing database and Data Lake.

It is an opportunity to further explore data and define the correct architecture that our clients need.



Possible outcomes

- Customer sentiment
- Foresee particular events such as the breaking of a machine
- Occupation of a given space
- Launch of possible market trends

Esplores: knowledge discovery platform

Esplores is a data visualization/interaction that creates virtual views and very effective analysis of any type of data – Big Data – whatever your operating field.

KEY POINTS

Reduces the complexity of Big Data projects.

Easily connects to multiple database types (Oracle, Microsoft, MongoDB, etc).

Integrated with Python, it implements Data Mining activities.

It helps companies in their digital transformation process, facilitating interactions between the ICT department and the Business Units.

esplores
the ultimate exploration tool

Big Data outputs

The use of the data owned by your company is the keystone of the whole process. The data itself can be retrieved from any source and analyzed to solve a variety of problems by obtaining:

Cost reduction

Process and/or decision time reduction

New product development and offer optimization

Data-driven decision

Connection between Big Data and Analytics, can have a strong impact on your business strategy whatever the area of relevance.

FOR EXAMPLES YOU CAN:

Real time identification of the problems and malfunctions reasons

Automated estimate of entire risk portfolios

Develop solutions and offers by monitoring customers' buying habits

Detect fraudulent behavior before it can hit you by putting your business at risk

How to use Big Data

Companies which are unable to use the opportunities Big Data give, will experience big difficulties in being innovative and they will also face a more ferocious competitiveness.

To create a successful business model, it is important to generate some functional flows whilst taking advantage of different data sources – this is clearly not an easy thing to do.

There are multiple use cases for every business model which relies on Big Data.

The real commercial value of such Big Data sources can be identified through specific applications which can be completely different from one another according to the sector in which they are used.

We can also identify some interesting technical challenges which arise with the integration and the management of all these data. Every single reality should understand the correct use case for their own enterprise before implementing it.

Many different companies are still unprepared and just starting to explore the potential use cases of Big Data.

AUTOMOTIVE

"In the next decades the vehicles will be characterized by several significant developments that will make it remarkably different from today. A fierce focus on innovation across the broad automotive landscape will be concentrated on software, electrical systems, electronics, engine and auxiliary systems, and powertrain."
Ben Stanley, Global Automotive Research Lead Institute for Business Value, IBM.

—

83% of people understand the real strategic value of data which can increment the operative efficiency, new business models, digital experiences for consumers, system integration, personalized services and the ability to connect to all aspects of a person's life. But only 18% says they are now working on a platform of digital data. [Source: Automotive 2030 studio "Racing towards a digital future"]

CYBER SECURITY

"By 2021, forecast costs for protecting and solving cyber attacks will be more than $1b. The product and services market is evaluated around $180b". These are the estimated made by Alessandro Profumo, CEO of Leonardo, the national group of aerospace and defence which has created a new branch just for cyber-security. [Source: Il Sole 24 Ore, 24th September, 2019]

—

IBM's Watson stores a massive corpus consisting of more than 10 billion elements (structured security information from security events and related data) and refreshes its understanding at the rate of 4 million more elements each hour. In a sense, Watson for Cyber Security is like a security expert who reads the web 24x7, never forgets, formulates hypotheses about attacks based upon this highly dynamic knowledge base and gets smarter over time.

HR-TECH

"The profession of the HR recruiter is swiftly changing due to the new technologies: Artificial Intelligence, Predictive Analytics, Gamification: these are all instruments which allow reducing the "time to hire" and improve precision in the selection process whilst eliminating prejudices and unconscious bias which influence our decisions". Pietro Novelli, HR recruitment specialist and Manager for the Italy branch at Oliver James Associates. [Source: Panorama, 17th October 2018]

—

HR Analytics are very useful tools which allow to monitor and manage in a data driven way critical business areas to support the choices made by the HR department, evaluate objectives and performances. Such tools are known as People Analytics or Talent Analytics and they analyze both the people and the process applications used for the human capital inside the company to improve performances and employees' loyalty.

ENERGY

In the next few years we will have access to a massive amount of data which will allow us to generate more reliable predictions whilst improving planning, management and saving energy. The production and the energy consumption of the next future will need to become smarter'. [Source: Il Sole 24 Ore, 25th June 2019]

—

By using mathematical models and having access to the list of consumption and forecast of hourly temperatures it is possible to have good simulations and forecast hourly and daily based consumptions.

FINTECH

Digital platforms allow to complete the process automations, from prospect client management to collection. Big Data can be used to gather information, do transactions online, implement automatic back office workflows. They can also support other factoring phases for new prospects such as small enterprises. They also allow to save money on operational costs. [Source: Il Sole 24 Ore, 24th September 2019]

—

Big Data in Fintech are a key tool for banks and, together with the use of AI, are one of the trends which is gaining more and more interest. With the growth of electronical records, financial services are actively using Big Data analysis to get in-depth analysis, archive data and create predictive analysis.

CASE STUDY / BIG TECH

Amazon

Founder: Jeff Bezos

Headquarter: Seattle

Employees: 647.500 (2018)

Revenue: 232,9 billion $ (2018)

Amazon has grown far beyond the online library of the early days, and in large part this is due to the extensive use of Big Data. One of its greatest innovations was the personalized recommendation system – which, of course, is built on the big data it gathers from its millions of customer transactions. At a higher level, its Amazon Web Services offers cloud-based computing and big data analysis on an enterprise scale.

Amazon has also incorporated big data analysis into its customer service operations due to their sophisticated Relationship Management Systems which made extensive use of their own customer data.

Amazon has grown far beyond its original inception and much of this is due to its enthusiastic adoption of big data principles. It looks set to continue breaking new ground in this field, for the foreseeable future.

CASE STUDY / BIG TECH

Google

Founder: Sergey Brin, Larry Page

Headquarter: Mountain View

Employees: 98.771 (2018)

Revenue: 136,81 billion $ (2018)

Google has not only significantly influenced the way we can now analyze big data (think MapReduce, BigQuery, etc.) – but they are probably more responsible than anyone else for making it part of our everyday lives. Google process 3.5 billion requests per day, and each request queries a database of 20 billion web pages.

A meaningful project is BigQuery, its commercial service for allowing companies to store and analyze Big Data sets on its cloud platforms.

Another big data project Google is working on is the self-driving car. Using and generating massive amounts of data from sensors, cameras, tracking devices and coupling this with on-board and realtime data analysis from Google Maps Streetview and other sources allows the Google car to safely drive on the roads without any input from a human driver.

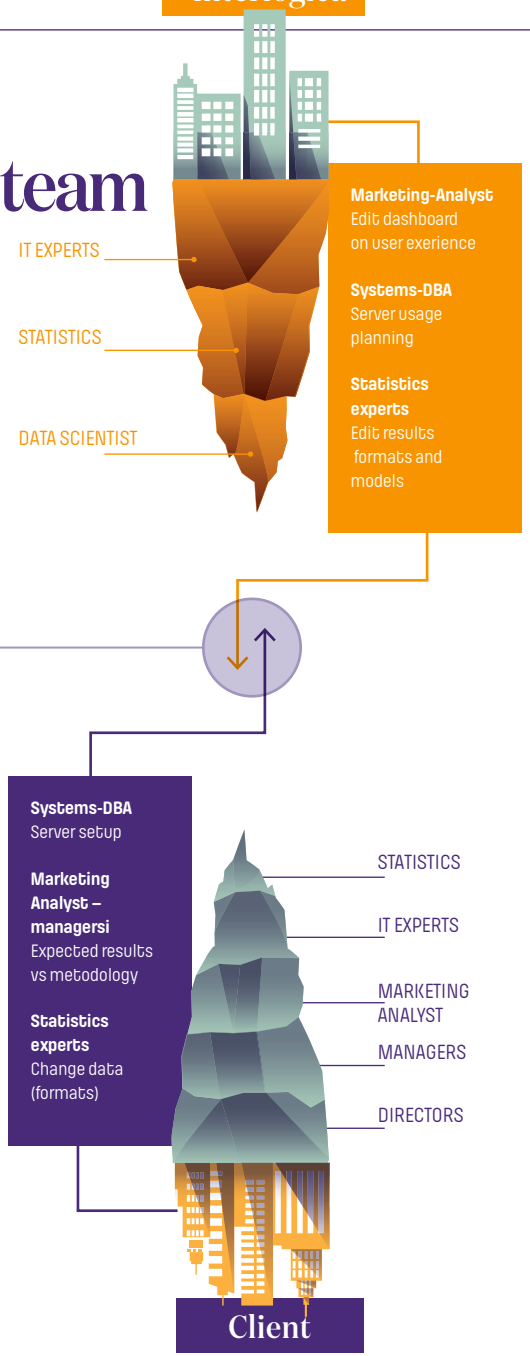
Data scientist and the dream team

A Data Scientist is an expert who knows how to understand and interpret data, which can be structured or unstructured. The goal is to reach determined business goals by transforming the data in new knowledge and opportunities.

As a professional he/she needs to have a clear vision of the most widespread working tools and of the most relevant platforms for every goal. This means that the scientist needs to know tools, typical statistical methods and Machine Learning. Most of the effort is to actually gather, clean, organize and format the data.

Whilst creating the Data Science team it is fundamental to pay attention to every single of one of its members. With all our experience, we suggest to have a multidisciplinary team in place as it is an optimal solution to face heterogenous projects.

We decided to have three different experts in our Dream Team: an IT expert who can easily design and optimize the software and manage the database; a Data Scientist who elaborates mathematical models to collect information from the data and designs solutions to organize them in a clear way for the user and a statistics expert who will carry out tests on models and will elaborate descriptive statistics which can easily explain the results of his work to the final user.



Periodical meetings

- POC discussion
- Work planning
- Progress
- Delivery
- After delivery

Data Science tools

The Data Scientist collects all the different data from different sources – such as business applications like ERP and CRM systems, data bank, web analytics, social media, electronic documents, IoT tools – and he/she structures them in organized and accessible formats.

In the scientist's toolbox there are softwares and programming languages to interact with the databases and extrapolate useful information (Data Mining). The data scientists can either elaborate algorithms or use statistical methods and Machine Learning tools following a scientific approach which is also result-driven.

Such process allows to spot in a big number of data, only the information which are valuable for the company, by building correlation models and explaining causalities whilst developing scenarios and predictive models.

A painful thing for many enterprises, especially SMEs is the cost of some softwares which own the data statistics and analysis. This has a huge impact on the budget. Luckily there are also some open source tools available on the market for different operating systems and they can compete with the other ones because of the developers' communities they have backing them.

Open Source is not just possible but is also highly encouraged.

INTERLOGICA

Lesson Learned

At Interlogica we create solutions for managing, analyzing and visualizing (structured and unstructured) Big Data from every existing database and Data Lake.

It is an opportunity to further explore data and define the correct architecture that our clients need.

Here are some important examples of how we effectively helped some clients to extrapolate value from the data they had, going through some processes, technologies and competencies in Analytics and Artificial Intelligence.

Big Data Analysis for the energy sector

GOAL

A platform for the estimate of clients switch out.

—

CONTEXT

A very important multi utility company in the energy sector decided to start an innovation project with us about Big Data through two main projects.

The first one was studying the switch out rate of its clients to maximize Customer Retention.

—

CHALLENGE

We had to understand the factors linked to the reasons why customers were switching to other companies through the analysis of the company's database, insight researches on their clients and by starting new commercial and marketing activities.

SOLUTION

We created a tool developed on Machine Learning techniques which allows to analyze a plethora of anonymous info associated with clients, such as personal data, the history of contracts and subscribed offers and the interactions the clients had with the company itself (complaints, info requests or marketing activities).

We used such data for every client to understand the possible risk of annulling the contract in the near future, in an automatized way. The multi-utility has then associated all the anonymous info to the real clients.

The Machine Learning algorithms we used are Random Forest which allow to determine decisional factors (for example the increase of energetic bills) so offers can be adjusted and developed for every single client. Also more correct estimates about cost/advantages to determine the right amount of investment which can be calculated.

GOAL

A platform for estimating electrical usage.

—

CONTEXT

Big Data and Artificial Intelligence have intertwined also in this second project which is about increasing the efficiency of the forecast process of the electric consumption.

—

CHALLENGE

Optimizing the buying process of the electrical energy whilst reducing the costs for the multi-utility and the energetic waste. Reducing the margin error in forecasting the consumption is a crucial point in this optimization process.

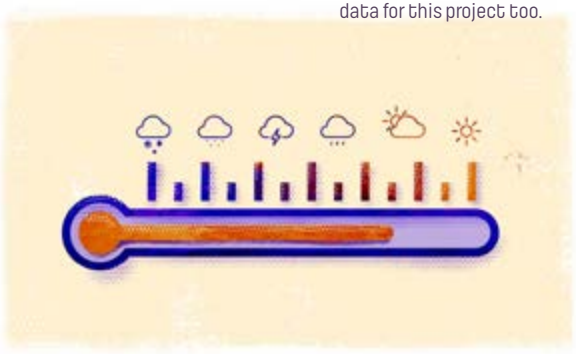
SOLUTION

We created a platform which analyses real-time electrical consumptions and weather forecasts: it combines the data with other info about consumers and calendar. These data then get structured in a database which uses the amount of data to train predictive models for electrical consumption.

These are decomposition models of historical data combined with Ensemble Learning methods automatically selected by the platform. It has been provided with a decisional model which allows to select the more adequate algorithm to foresee the electrical consumption about the type of consumer and the period of the year in which the forecast need to be made.

To allow the client to interact in an agile way with the data and to visualise the consumptions forecast by the platform we gave them an interactive dashboard realised with the help of Esplores.

Models have been elaborated with anonymized data for this project too.



Big Data for legal invoicing

GOAL

A platform to manage the flow of electronic invoices.

—

CONTEXT

One of the biggest players in Italy for certification authority has started some investments in innovation projects using Big Data so the flow of electronic invoices of clients can be managed (the system is designed to work supporting 500 invoices per second).

—

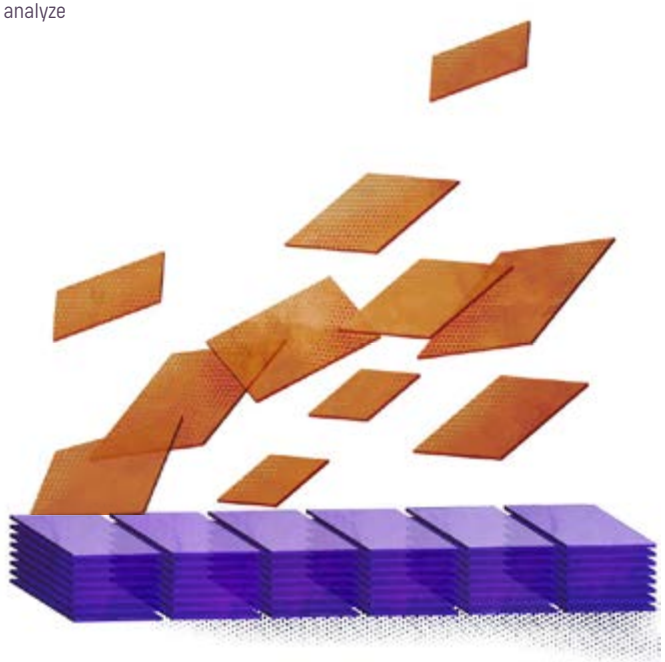
CHALLENGE

To create a flow of automatic management for invoicing which allows to archive and analyze massive amount of data.

SOLUTION

After a careful evaluation of the needs and the expectations of clients, we decided to use a technology called Apache Kafka combined with a Hadoop cluster specifically designed to archive and analyze big quantities of non structured data in a distributed calculation environment.

All these sets of data are divided into an Oracle database which allows a real time elaboration of metadata. The sets of data will remain there for three years as we estimated the analysis of almost a petabyte of data will be carried out in such timeframe.



Natural Language Processing applied to automatic email dispatching

GOAL

Automatic sorting of PEC (Certified) emails.

—

CONTEXT

A fourth project in which we were involved was developing an Artificial Intelligence platform to automatize the sorting of certified emails for a Financial Services company.

—

CHALLENGE

Improving, quickening and automatizing the analysis functions of the text to identify the emails recipient in an automated way.

—

SOLUTION

We decided to use the Natural Language Processing (NLP) method, a computer processing of the natural language paired with Machine Learning techniques.

For such project our Data Science team has analyzed a cluster of mails to build a classification model of the text.

These models allow to classify the incoming emails according to their context (sender, subject, body) automatically forwarding them to the right group of work or recipients.

The platform allows to create automated emails to reply on the context of the email received.

The client also has the chance of correcting some inaccuracies that might happen in the automatic sorting communicating the right recipient to the platform; the algorithms of the Artificial Intelligence modify themselves automatically according to the new info, incrementing the precision of the platform itself.

The result is an accurate and trustworthy categorization of emails which translates into less mistakes and time wasted from an employee perspective.

To summarise:

- It uses NLP AI technology.
- It automatically classifies emails as soon as they come in.
- It generates an automated response according to the correct context of the email.
- It reduces mistakes and helps to save time for employees.

Vocabulary

PREDICTIVE ANALISYS

Use of data, statistical algorithms and Machine Learning techniques to identify the probability of future results based on historical data..

—

BIG DATA

Big Data is the expression used to identify a very consistent set of data, to the point of making it difficult to manage with the usual tools. Generally Big Data meets the 3V rule: Volume, Variety, Velocity.

—

CLUSTER

Group of servers that allow to ensure continuity of service and distribution of calculation / network load.

—

DATA ANALYSIS

A process of data inspection, cleaning, transformation and modeling with the aim of highlighting information that supports strategic business decisions.

—

DATA ANALYST

Expert who collects, processes and performs statistical analyses transforming data into a series of information to be communicated to company management.

—

DATABASE

Organized collection of data, generally stored and accessed electronically from a computer system.

—

DATA ENGINEER

Expert typically in charge of managing data workflows, pipelines, and ETL processes that prepare and transform data for data scientists.

—

NON STRUCTURED DATA

Information that either does not have a pre-defined data model or is not organized in a pre-defined manner (texts, emails, videos, messages on social media).

—

DATA SCIENCE

Set of methodological principles and multidisciplinary techniques aimed at interpreting and extracting knowledge from data.

—

DATA SCIENTIST

Expert in charge of collecting and organizing large amounts of data, and extrapolating insight, analysis and reports.

—

STRUCTURED DATA

Standardized format for providing information about a page and classifying the page content (website traffic data, bank details or GPS coordinates collected by a smartphone).

—

DATA WHAREHOUSE

Collection or aggregation of structured data, coming from database management systems (DBMS) internal and / or external to the corporate information system.

—

EXTRACT TRANSFORM LOAD (ETL)

Data extraction, transformation and loading process in a synthesis system (data warehouse, data mart, etc).

—

HADOOP

Application framework distributed by Apache Software Foundation under Apache license, intended to handle large volumes of data.

—

ARTIFICIAL INTELLIGECE (AI)

Set of methodologies and techniques to reproduce or emulate functions and reasoning typical of the human mind.

—

MACHINE LEARNING (ML)

AI sector that deals with learning structure or behavior from data, to emulate them autonomously and automatically.

—

OPEN DATA

Data that is freely accessible: it can be used by anyone without copyright restrictions, patents or other control mechanisms.

—

PYTHON

Object-oriented programming language usable for many types of software development. Distributed with Open Source license.

—

R

Programming language and a specific development environment for statistical data analysis. Distributed with Open Source license..

—

REPLICATION

Process of copying files to different points of an architecture to improve reliability and fault tolerance.

—

SCALA

General-purpose multi-paradigm programming language designed to integrate the features and functionalities of object-oriented languages and functional languages..

—

STRUCTURED QUERY LANGUAGE (SQL)

Language used to interact with databases, for example to extract information..

Contact an Expert

To learn more about how Interlogica can help you unlock the value created through the Big Data analisys.

Follow us

in company/Interlogica
@Interlogica
@Interlogica

Further information

interlogica.it

Certifications

CSQ ISO 9001
CSQ ISO/IEC 27001

Contacts

Venice – Mestre
Via Miranese, 91/4 – 30174, Mestre (VE)
+39 041 5354800
info@interlogica.it

Milan
Via Francesco Caracciolo, 26 – 20155 Milan
+39 041 5354802
info@interlogica.it

Dubai
Dubai Silicon Oasis, Head Quarter Building Office
PO Box: 71245 – Dubai, United Arab Emirates
alex.jacobs@interlogica.it

Coding the future

interlogica.it