Hongo amanita muscaria - San Antonio de Pereira (Antioquia)
Conocido como matamoscas o falsa aronja

State of the Art: Big Data and Business Intelligence

SYNOPSIS:
This article presents the state of the art of Big Data and Business Intelligence based upon presentations made at a September 2016 congress held at the Catholic University of Pereira and the knowledge of the authors from implementing this type of system in transactional and decision support systems.

SÍNTESIS:
Este artículo presenta el estado del arte de Big data y la inteligencia empresarial a partir de ponencia presentada en el encuentro realizado en septiembre de 2016 en la Universidad Católica de Pereira y el conocimiento de los autores sobre la implantación de este tipo de sistemas en el soporte de decisiones transaccional.

ABSTRACT:
Managers must improve their knowledge of data resources and how to use them in their companies, especially in PYMES (small and medium-sized companies). In many companies, managers believe that data-driven solutions are only for managers in large companies because the cost is very expensive. In some companies, the information systems are antiquated and in others it has not been politically acceptable to save the data generated. These are problems in companies, especially in developing parts of the World. For example, in Colombia managers are still developing data-driven cultures and developing information systems that can help their firms compete with other companies in other markets. This article presents the state of the art of Big Data and Business Intelligence based upon presentations made at a September 2016 congress held at the Cahtolic University of Pereira.

RESUMEN:
Los directivos deben mejorar su conocimiento de los recursos de datos y cómo utilizarlos en sus empresas, especialmente en las PYMES (pequeñas y medianas empresas). En muchas empresas, los administradores creen que las soluciones basadas en datos son sólo para los administradores de las grandes empresas, porque el costo es muy elevado. En algunas empresas, los sistemas de información son anticuados y en otros no han contado con políticas aceptable para guardar los datos generados. Estos son problemas en las empresas, especialmente en los países en vías de desarrollo. Por ejemplo, en Colombia los gerentes siguen desarrollando culturas basadas en datos y desarrollando sistemas de información que pueden ayudar a sus firmas a competir con otras compañías en otros mercados. Este artículo presenta el estado del arte en Big Data y la inteligencia de negocio, sustentado en la presentación realizada en el encuentro celebrado en Septiembre de 2016 en la Universidad Católica de Pereira.

KEY WORDS:
BI, Business Intelligence, Big Data, DSS, Decision Support Systems, Transactional Systems, ICT Information and Communication Technologies, Mincit Ministerio de Ciencia y Tecnología

JEL CLASSIFICATION: C8, D23, D78, F23, H11, L86

PALABRAS CLAVE:
BI, inteligencia de negocios, Big Data, DSS, sistemas de apoyo a la decisión, sistemas transaccionales, tecnologías de información y comunicación TIC, Ministerio de ciencia y tecnología

CLASIFICACIÓN GEL: C8, D23, D78, F23, H11, L86

State of the Art: Big Data and Business Intelligence[1]

Andrés Henao Rosero[2]
Daniel J. Power[3]

"Nowadays computers are in the middle of most economic transactions. These "computer-mediated transactions" generate huge amounts of data, and new tools can be used to manipulate and analyze this data."

Hal R. Varian

Estado del arte: Big Data e inteligencia empresarial

Primera versión recibida el 8 de Septiembre de 2017, Versión aprobada el 16 de Octubre de 2017

Para citar este artículo: Henao Rosero, Andrés (2017). State of the Art: Big Data and Business Intelligence. En: Revista Gestión y Región N° 24 (Julio-Diciembre de 2017); pp. 9-22

INTRODUCTION

Today the companies are creating a data-driven digital world in which every day the amount of data is increasing at what seems to be an exponential rate. Today all companies have different types of information systems, this was a strategic moment that differenced with other companies, but today is not a difference, the difference is how the data and the information is utilized like knowledge, this for better decisions.

In the beginning of the computing age, the idea of information systems was to improve processes, making them faster, more reliable, more accurate and in some cases safer to use, but these systems generated and required more and more data. For many decades, the data was generated, but it was costly to store and challenging to analyze. The amount of data captured and stored has gotten larger and larger because the cost of capturing and storing data has gotten less and less expensive.

Computing began with the Abacus, the counting frame. This technology innovation was a manual calculator and the innovation improved the speed and accuracy of calculations. Other manual machines were also developed like the Slide Rule developed in 1620 by Pascal, a wheel used for efficient multiplication and division developed in 1694 and punched cards developed

1 *Presentations made at a September 2016 congress held at the Catolica de Pereira University and the knowledge of the authors from implementing this type of system in transactional and decision support systems.*

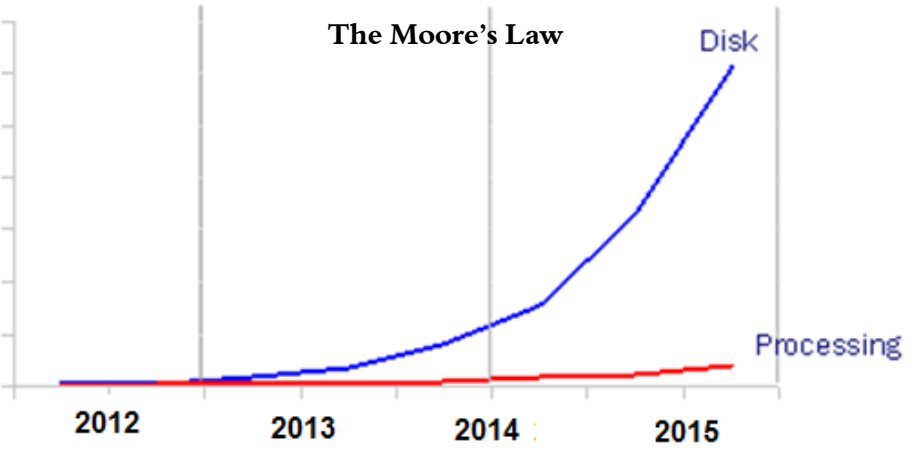
2 *Ingeniero de sistemas. Especialista y Maestría en Administración de empresas. Especialista en Gerencia de Proyectos de la Universidad de EAFIT. Decano de Faculta de Ciencias Económicas y Administrativas de la Universidad Católica de Pereira*

3 *Administrador de Negocios. Master of Business Administration y Doctor en Administradcion de Negocios de la Universidad de Wisconsin. Profesor de University of Northern Iowa*

for controlling the operation of looms. The Babbage Difference Engine was created to automatically calculate polynomials. It was a dedicated, single task machine. In 1890, the U.S. Census Bureau used 56 million punched cards to store data and provide input to tabulating machines. Alan Turing developed the concept of a general computer to process any algorithm, a Turing Machine in 1936. Modern computing traces its roots back centuries, but development of electronic computers started to accelerate in the 1930s and 40s (see http://www.computerhistory.org/timeline).

Today companies have three different type of systems, 1) decision support, 2) office support, and 3) transactional systems. The most common office support systems are word processors, spreadsheets, slide presenters and email. In some cases, desktop database systems like MS Access are office support systems. Transactional systems insure transactions are atomic, consistent, isolated, and durable. Many companies have sophisticated accounting or financial systems, and in some cases Enterprise Resource Planning (ERP) systems. Companies today also have websites with home pages and fan pages. Many companies do not have specific policies to administer web activities and social networks.

Today most, if not all, companies can´t analyze the vast amount of data that is produced every day by internal systems. Each year the gap between the data produced and the capacity to analyze the data is growing bigger. Moore's law (1965) predicts smaller and faster processors yet comparing data growth and the capacity to analyze the data, the gap between data growth and processing power is widening, this is showed in the Moore's Law.



Ilustration 1: Data and Processing Gap the Moore's Law

In 1989, the term of Business Intelligence (BI) was introduced by Howard Dresner (after this an analyst of Gartner Group) propose the idea that BI more popular the end of 90's). The idea was to create a specialized database called a data warehouse or data mart with consolidated data optimized for aggregations and queries. When this technology was initially adopted the cost of implementation was very expensive for both hardware and software and was

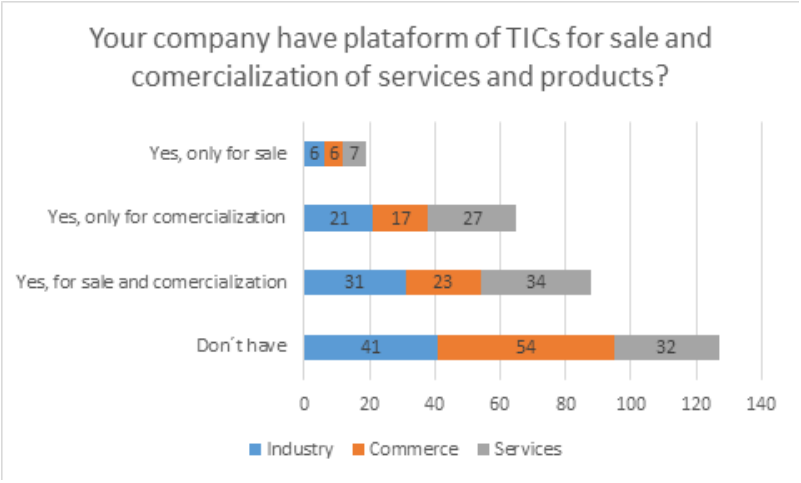
only used by large companies. In the early years, the principal problem associated with implementing the new BI software and technology was the quality of data. That problem persists today. In recent years, the increasing volume and variety of data, big data, has made business intelligence and decision support even more challenging.

Data from the World Economic Forum (WEF) published in 2016, indicates the use of Information and Computing Technologies (ICT) in Colombia is very low. According to this data, the country was in 68th position of 139 countries and fell four positions in the world ranking with the penetration as the prior year. This ranking is based upon the following factors:

- Political and government environment
- Business and innovation environment
- Infrastructure
- Accessibility
- Ability
- Use by individuals
- Business use
- Government use
- Economic impact
- Social Impact

Based upon these factors, the adoption of ICT in Colombia continued to grow in the individual use, internet, cellular phones, and PCs, but in companies, and the government growth in the use of ICT has been slow.

Based on the above, the Gran Ecuesta Pyme (GEP) carried out a survey:



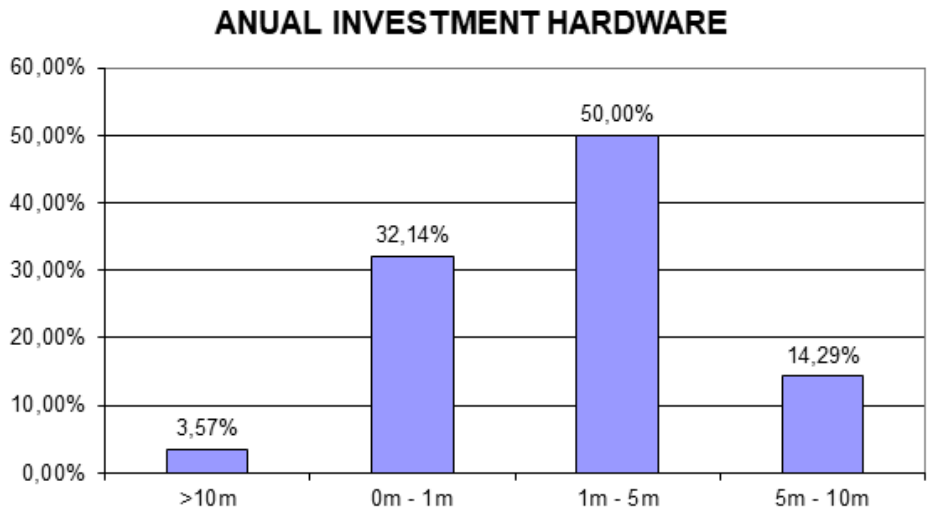
Source (ANIF, 2016)

Ilustration 2: ICT usage

In Figure 2, is possible view that the principal sectors of the economy, of the companies surveyed, 41% industry, 54% commerce, 32% services need tools. Only 31% of companies in the industry sector, 23% commerce, 34% services said that they have tools for buying and selling products and services. And less than 10% use tools only for selling (6% in industry and commerce sector, 7% in services).

In the Eje Cafetero region, the use of ICT in PYMES is not different that the national statistics. A study of analysis of utilization of ICT in the companies in Pereira City (Henao, Isaza, 2013) supports this assertion. In this study, the focus was on the construction sector and the results were very similar to the statistics reviewed with the survey of ANIF in 2016.

The use of ICT in these types of companies and sectors is very low because it is not normal that these types of companies have the needed level of investment in hardware (see Figure 3).

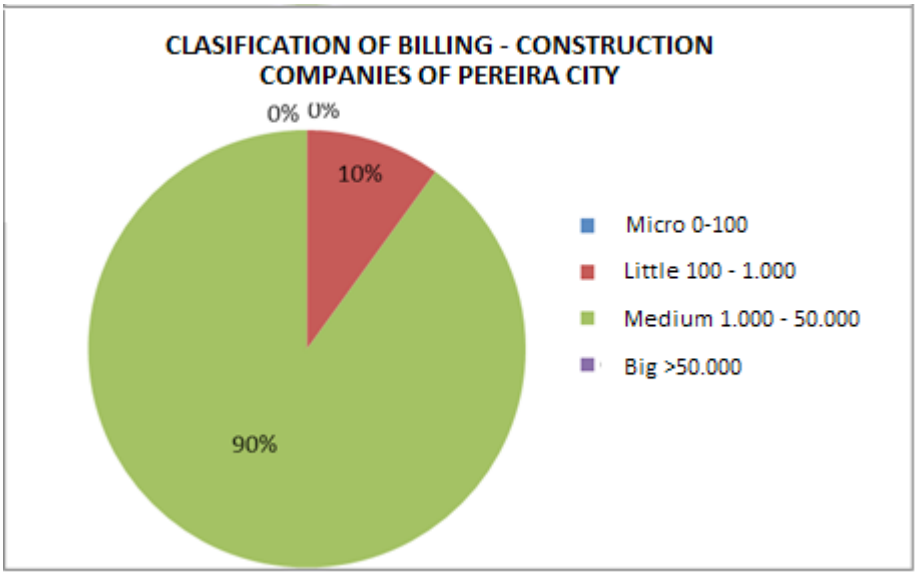


Source: Henao, Isaza 2013

Ilustration 3: Investment in Hardware

Source: Henao, Isaza 2013. (m, million of Colombian pesos)

The above figure shows the level of investment in computing hardware by construction companies in Pereira City. The investment in physical infrastructure is between 1millions – 5 millions of Colombian pesos for 50% of the companies and 14,29% between 5m – 10m of Colombian pesos, 32,14 % between 0m – 1m of Colombian pesos and only 3,57 higher than 10 million.



Source: Henao, Isaza 2013

Ilustration 4: Billing

The billing of the companies is between 100 – 50.000 millions Colombian pesos.

According to this study, if these sectors have more participation in creating the gross domestic product (**Produsul intern brut** PIB) of Colombia, there is higher probability that other sectors have the same statistics or lower, and very similar to the statistics at the national level in the survey of ANIF and WEF, 2016.

The country has a low utilization of ICT, the priority is to find ways to implement BI technology and Big Data technology. This article is a start at achieving that goal.

State of the Art

In recent years, managers have realized that information systems are not a cost for companies, rather these systems are more like an investment to realize a strategic growth plan. Today the systems are less costly for the capabilities purchased. The cost of both the hardware and software are cheaper, the Internet growth has led to more online transactions in the systems. Prior to the Internet transactions were limited to internal systems.

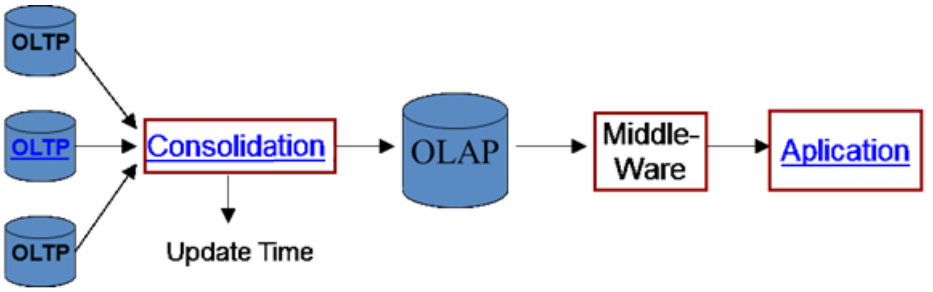
With the Internet a user can interact with companies via systems on the web. The number of transactions has increased, but the capability to generate reports is more difficult. The Database Management Systems (DBMS) were more for

transactional systems, with the principal functions of insert, update, modify and delete records, but not for reports that companies and departments needed for month to month or weekly reporting. When these types of reports were executed, the performance of transactional systems was degraded significantly. Also, analysts in departments in a company executed the reports at different time, data changed and sometimes different databases were used. Therefore, it was common to have inconsistencies and problems with the statistics when these reports were presented to the company president or other managers.

These problems led to Business Intelligence systems and more professional processes to collect, process and distribute information for decision support.

These systems have hardware and software tools with special capabilities. Normally a data mart or data warehouse system is a special purpose database system that is separate from transactional systems.

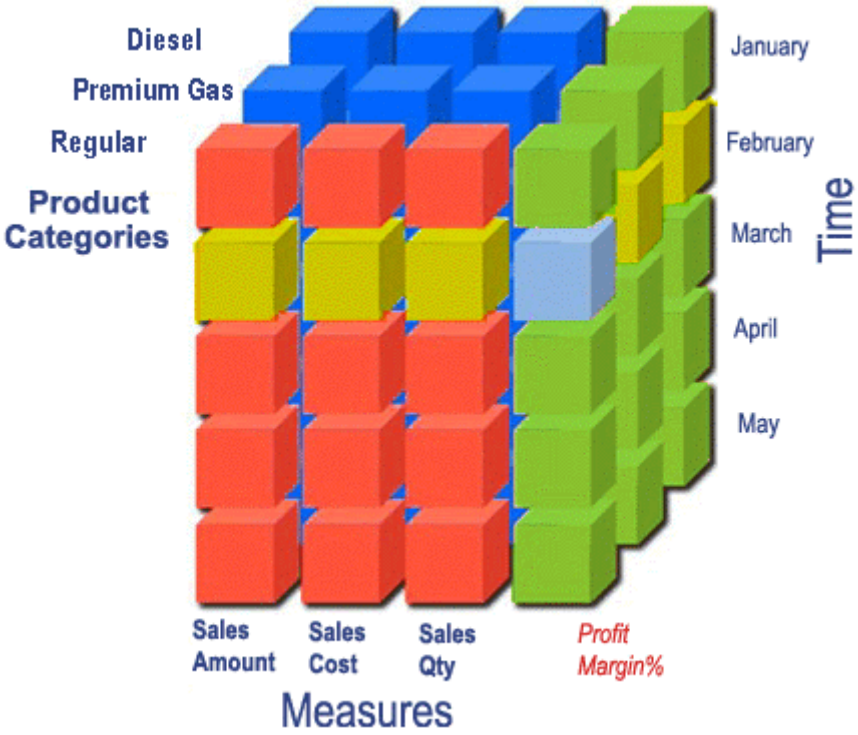
Because of this separation, companies must invest more in hardware and software to create the architecture in Figure 5.



Own Source

Ilustration 5: Business Intelligence architecture

In this architecture, OLTP is the transactional system that updates periodically a data warehouse. OLAP or Online Analytical Processing is a data cube with dimensions and a fact table (see Figure 6). Creating appropriate dimensions and identifying appropriate data is the major goal of a business intelligence effort. While facts correspond to events, dimensions correspond to people, items, or other objects. For example, in the retail scenario, purchases, and returns are facts. On the other hand, customers, employees, items and stores are dimensions and should be contained in dimension tables, cf., Chapple (2017).



Source http://www.sql-datatools.com

Ilustration 6: Data Cube

Today are many free tools for the middleware of the data warehouse and Relational Database Management Systems (RDBMS). The hardware is low cost, and it is very possible that small companies in Columbia like PYMES can implement this type of technology, but managers must make a special effort to insure the quality of data. Many small companies in Columbia don´t have a special purpose transaction system, and much data is managed by Excel. This creates the possibility that the integrity of this data is not good. In a Business Intelligence system, the quality of data is very important to the success of implementation. The figure 7 identifies characteristics and benefits of quality of data.

Quality Characteristics	Benefits
Correct Data	meets Information need
Adequate completeness	All information is provided
Appropriate context	Known meaning
Needed precision	Can trust the information
Good format	Can use the information easily
Timely	Provided when needed
Right place	Where needed
Appropriate propose	Reach objectives and satisfy client

Ilustration 7: Quality Characteristics

Data is increasing and some of it can better inform decision making. The Gartner Hype Cycles 2012 – 2014 (see Power 2016) show:

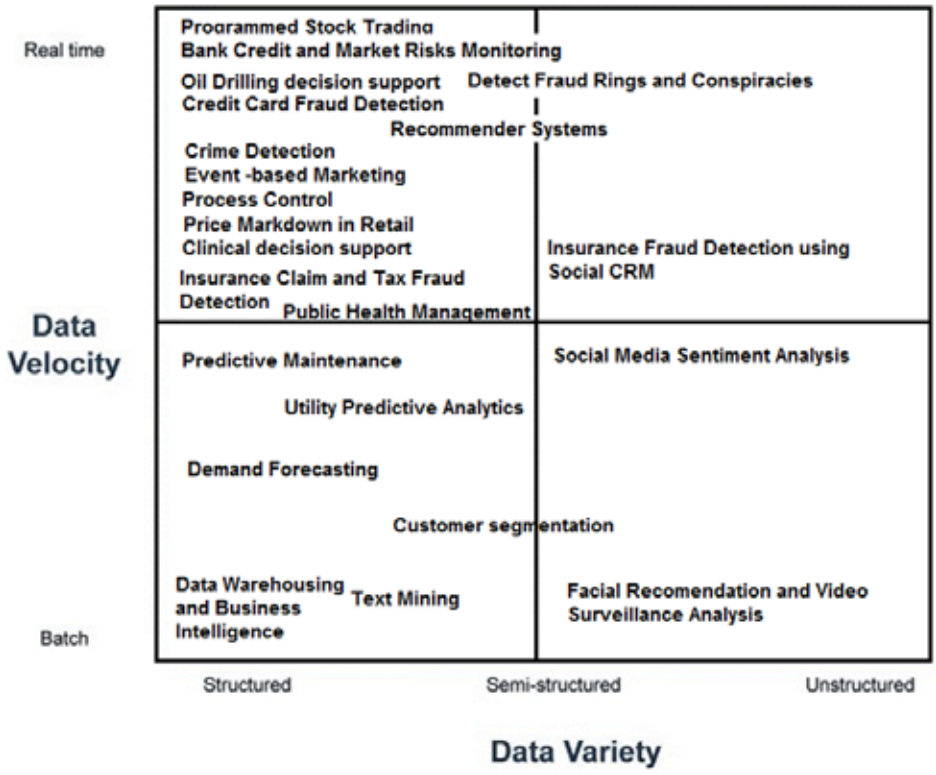
- In July 2012, noSQL Databases and Data Science were rising technology trends, while Social analytics was at the peak of inflated expectations. Predictive analytics was at the plateau of productivity.

- In July 2013, Big Data was at the peak of inflated expectations. Content analytics was moving toward the trough of disillusionment. Predictive analytics was at the plateau of productivity. Prescriptive analytics was a rising technology trend.

- In July 2014, Big Data and Content Analytics was moving toward the trough of disillusionment and the Internet of Things was at the peak of inflated expectations. Data science and prescriptive analytics were rising technology trends. Virtual reality and In-Memory analytics were on the slope of enlightenment. Big data is at the marketing stage and it is a marketing term.

In 2001, Doug Laney at Gartner describes big data in terms of Volume, Velocity and Variety, and in 2012 Gartner updated the definitions. "Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."

Big data can be used in many analytics decision support use cases. Figure 8 from Power (2016) shows different uses for differing Data Velocity and Data Variety.



Ilustration 8: Uses of Big Data

The use of Big Data is expanding in many companies and sectors, below are some examples:

- Customer analytics
- Data-driven products and services, including personalization systems
- Enterprise Data Warehouse optimization and data warehouse modernization
- Operational analytics and monitoring business operations, and
- Fraud detection, compliance, security/intelligence extensions

In telecommunications companies:

- Customer Experience Management: Targeted Marketing & Personalization, Predictive Churn Analytics, Customer Journey Analytics , Proactive Care
- Network optimization: Network Capacity Planning & Optimization, Network Investment Planning, Real-Time Network Analytics
- Operational analytics: Revenue Leakage and Assurance, Cyber Security and Information Management, Order-to Activation, Customer Care Optimization

- Data monetization: Data Analytics as a Service, IoT and M2M Analytics, New Revenue sources – Connected Home, Connected Car, etc.

Intelligence Industry Survey (2014)

Insurance companies are starting to collect data on driving habits utilizing sensors in their customers' cars. Sun, Heller (2012).

In Logistics, implemented a powerful intelligence solution that could identify trends in demand and generate a rolling sales forecast.

Retail opportunities:

- Personalized recommendations
- Dynamic Pricing
- Store Experience
- Micro segmentation and inventory management

Safety

- Provides law enforcement agencies with actionable intelligence by aggregating massive amounts of data to visualize patterns and trends for predictive policing and crime reduction efforts. Haisler (2014)

Health Care and Life Sciences

- Approximately 80% of data is unstructured and the volume is increasing exponentially
- Output from medical devices, doctor's notes, lab results, imaging reports, medical correspondence, clinical data, and financial data
- For example, a large Healthcare organization that serves more than 100 million members collects many petabytes of data

Video Calling

- Real-time recognition of faces, objects, events, and then create alerts
- Archival analysis of video to investigate activities
- Quality control, security, policing, customer service

See Sondergaard (2017), Gartner Research "The information is the Oil of the 21st century and analytics is the combustion engine"

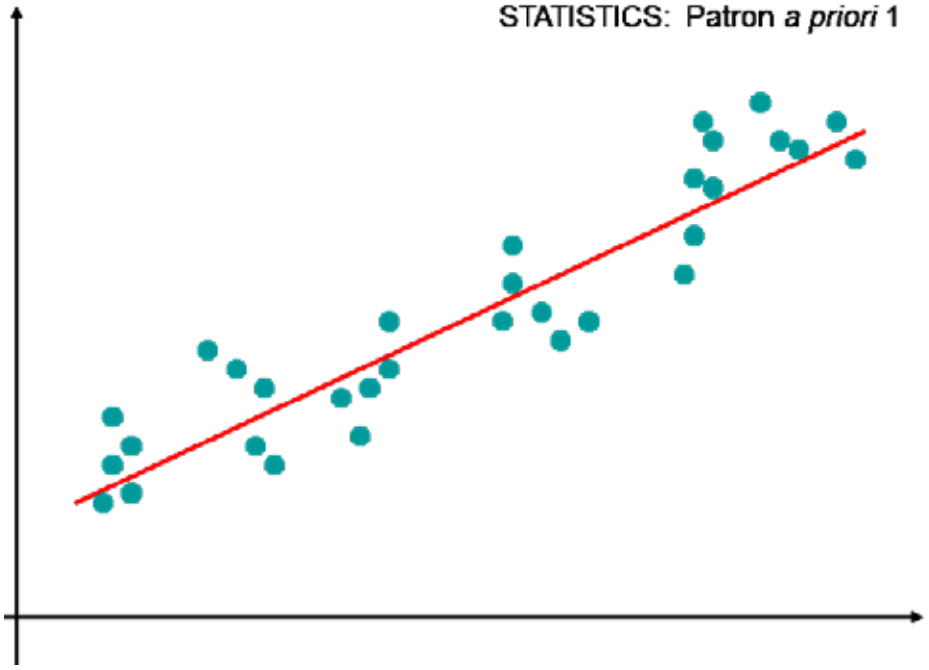
Today companies have a many problems and possibilities that may be resolved with Big Data. New data sources can help answer questions that people have

when making decisions. For example, data patterns about purchase activity or website visits might help understand a customer (see Figure 9)

Ilustration 9: Data pattern Garcia,(2016)

The data pattern must be analyzed in the context of a conceptual framework. Perhaps the customer in Figure 8 is visiting the website more and is visiting more web pages on each visit.

Customer data is often analyzed in different ways that create useful information. Perhaps a regression analysis shows a trend line in the customer's behavior (Figure 10).



Ilustration 10: Trend line Garcia (2016)

Today there exist different ways to exploit data like induction algorithms, data mining, auto organized maps, and bayesian networks. With this type of technology, it is possible that Columbia PYMES can implement them and differentiate themselves from other companies and better compete with companies from other countries with high standards of quality and high levels of technology use.

Big Data has the five V´s, Volume, Variety, Value, Veracity, Velocity, Colombia PYMES data has all five characteristics, but the Veracity or truthfulness characteristic is a difficult in some PYMES.

For example, social networks have generated a lot of data, but not all the data is trustworthy. A few years ago another new term was first used for the generation of data by machines and devices that is the Internet of Things (IoT). Smart TVs, smart watches, even clothes can generate and capture behaviors of people that is important to analyze. Improving transactional systems seems a very low priority in Columbia companies and the problem of data quality from these sources remains challenging.

CONCLUSION

- Big data is a reality, but the claims of vendors about its uses often seem over stated.
- Analyzing "big data" to find a great business plan or to identify the next revolutionary product idea seems like wishful thinking.
- Evaluating potential operating decision-making use cases can help managers decide if resources should be dedicated to exploiting "big data".
- Researchers should develop use case diagrams that show how an analytical task is performed using a new data source with one or more of the "big data' characteristics of high volume, high variety, high velocity, high volatility and high variability.
- There should be optimism about finding novel use cases for new data sources.
- Better defined decision-making use cases can help senior managers assess the value of new data sources.
- Identifying and documenting decision support use cases for new and expanding data sources is an urgent need
- Managers perceive and anticipate better outcomes from decisions if more facts are used in decision making
- Analytics, communications, data technologies and decision support are enabling a global, data-driven society
- Our emerging global society is highly interconnected and many people rely on information technology to support decision making
- The increasing volume, velocity and variety of data is important to building new decision support functionality

- The quality of data is very important, special in Columbia PYMES that cannot implement transactional systems and invest in ICT.
- It is a priority that Columbia companies increase the use of ICT. Today the cost of hardware and software is decreasing, and the Columbia government has multiple projects to help PYMES implement better ICT, the first step to implement Big Data and Business Intelligence systems.

BIBLIOGRAPHY

- Computer History Museum (2017), Computer History Timeline. Available: http://www.computerhistory.org/timeline/
- ANIF, (2016). Cómo va la implementación de las Tecnologías de la Información y Comunicación (TICs) en las Pymes de Colombia?, Available: http://anif.co/sites/default/files/archivosgenerales/actualidadpyme97.pdf
- WEF, (2016). Informe Global de Tecnología de la información 2016. Available: http://www.cdi.org.pe/InformeGlobaldeInformacion/GITR2016.html
- Zorzi, (2011). Las TIC en el desarrollo de la pyme. Available: http://pymespracticas.typepad.com/files/tic-y-pymes-en-al-final-2011.pdf
- Isaza, Henao (2013), Análisis del grado de utilización de las TIC en las empresas constructoras de la ciudad de Pereira.
- Power (2005). Decision Support Systems FAQ.
- Zikopoulos , Deroos , Parasuraman, Deutsch, Corrgigan, Giles (2013), Harness the Power of Big Data.
- Power (2013), Strategic Planning Concepts.
- Varian (2014), Big Data: New Tricks for Econometrics.
- Chen, Chiang, Storey (2012), Business Intelligence and Analytics: From Big Data to Big Impact.
- Chapple (2017), Facts vs. Dimensions Tables In Database Understanding Key Business Intelligence Terms, Computer Science
- Telecoms.com (2014), Intelligence Industry Survey. Available www.cloudera.com/content/dam/cloudera/Resources/PDF/solution-briefs/Industry-Brief-Big-Data-Use-Cases-for-Telcos.pdf
- Sun,H, P. Heller (2012), Oracle Information Architecture: An Architect's Guide to Big Data.
- http://www.ibmbigdatahub.com/presentation/big-data-and-analytics-manufacturing-industry-vaasan-group
- Haisler (2014), Top 10 Uses of Predictive Technology Making Our Communities Safer Government Technology. Available: www.govtech.com/photos/Top-10-Uses-of-Predictive-Technology-Making-Our-Communities-Safer.html
- www.mapr.com/solutions/industry
- ARNOTT, D.; PERVAN, G. A critical analysis of decision support systems research. Journal of Information Technology, v. 20, n. 2, p. 67–87, 12 abr. 2005.

- ARNOTT, D.; PERVAN, G. A critical analysis of decision support systems research revisited: the rise of design science. Journal of Information Technology, v. 29, p. 269–293, 2014.
- FROLICK, M. N.; ARIYACHANDRA, T. R. Business Performance Management: One Truth. Information Systems Management, v. 23, n. 1, p. 41–48, dez. 2006.
- MARCH, S. T.; HEVNER, A. R. Integrated decision support systems: A data warehousing perspective. Decision Support Systems, v. 43, n. 3, p. 1031–1043, abr. 2007.
- MCAFEE, A.; BRYNJOLFSSON, E. Big data: the management revolution. Harvard business review, n. October 2012, p. 60–68, 2012.
- OLSZAK, C. M.; ZIEMBA, E. Business Intelligence Systems in the Holistic Infrastructure Development Supporting Decision-Making in Organisations Business Intelligence Systems in Decision-Making. Interdisciplinary Journal of Information, Knowledge, and Management, v. 1, 2006.
- RANJAN, J. Business intelligence: Concepts, components, techniques and benefits. Journal of Theoretical and Applied Information …, v. 9, n. 1, p. 60–70, 2009.