

A well known traditional saying

"Information is Power"

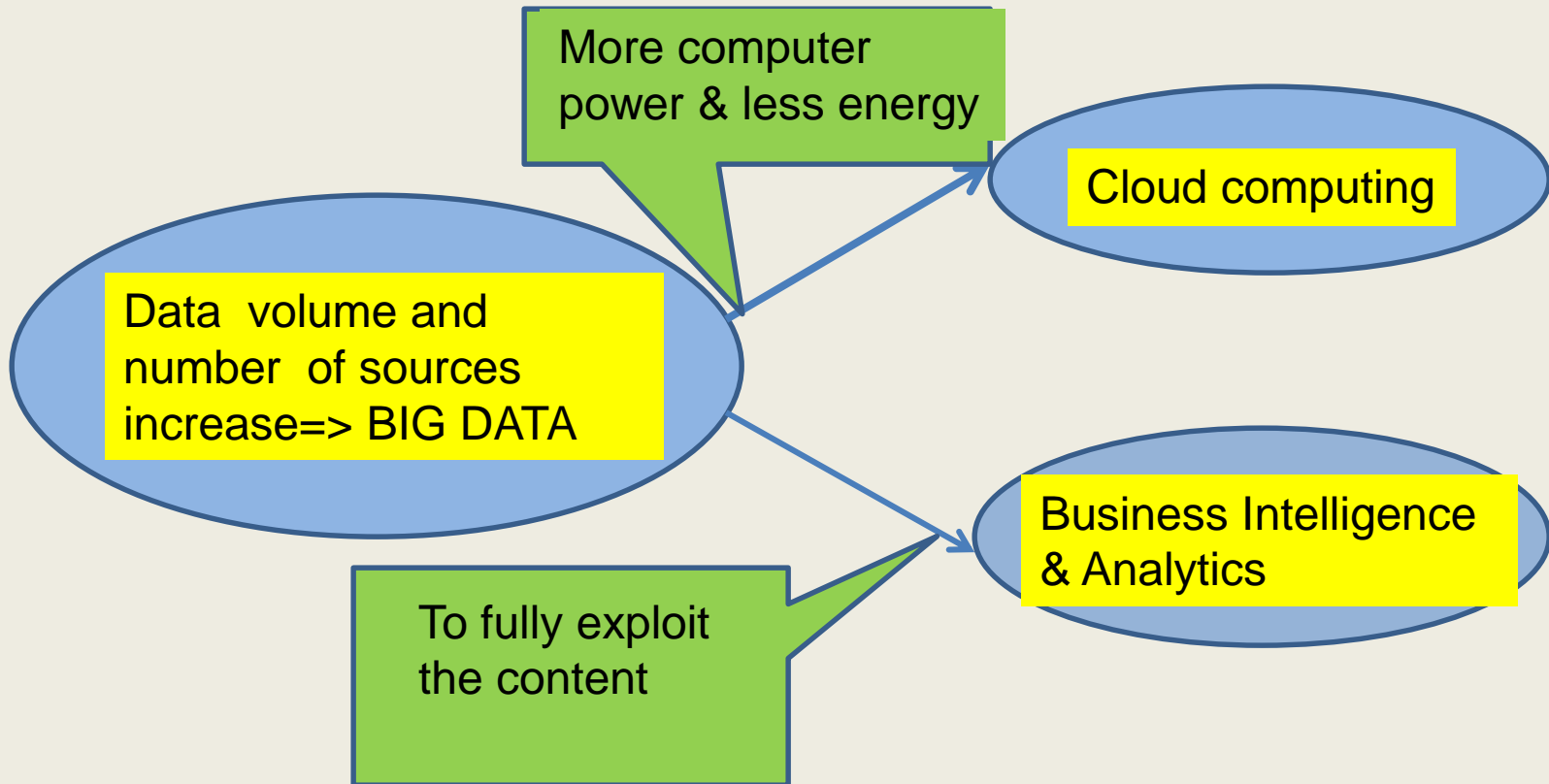
At present,

"Information is Power"

is replaced

"Knowledge is Power"

Two Key Enabling Technologies in Europe



BIG DATA in ROMANIA: CONTEXT and CURRENT STATUS

by F. G. Filip

The Romanian Academy, Bucharest:
the Library, and INCE

This talk is about

modern Business Intelligence & Analytics
technologies

and

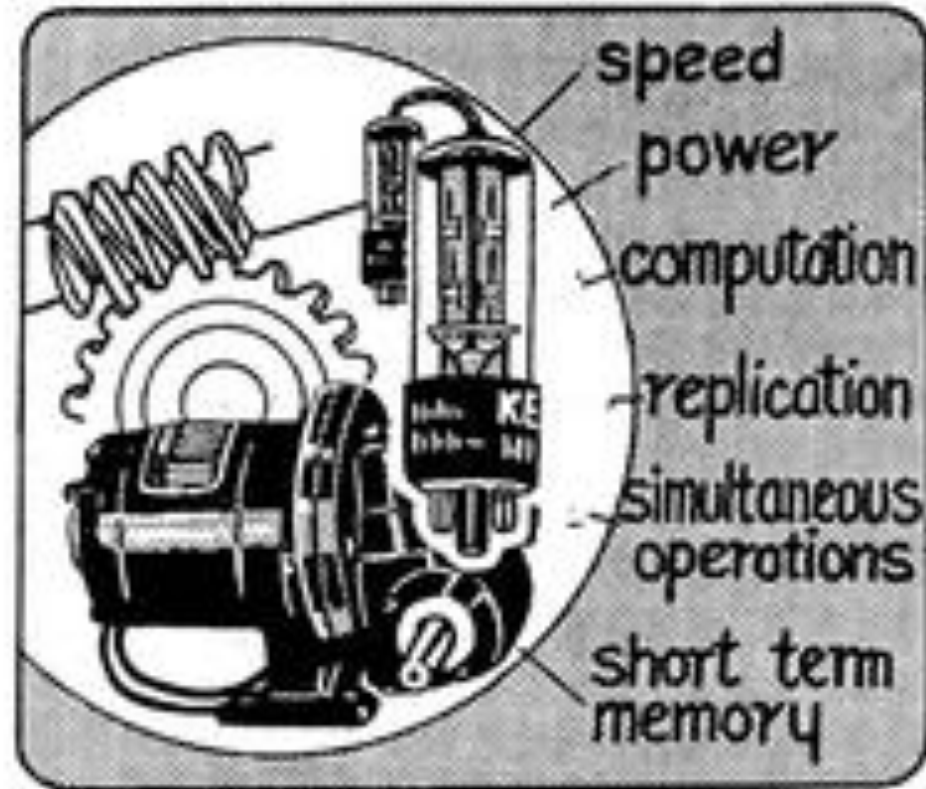
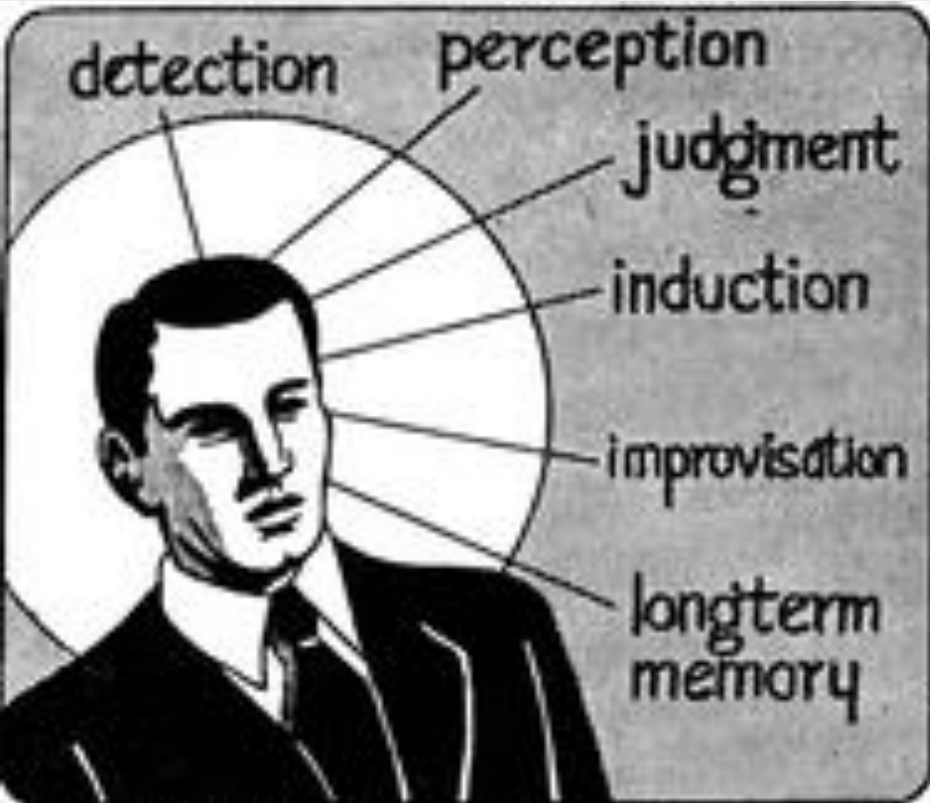
their usage in processing big amount of
data in management and control
applications

– It gives a brief account of BIG DATA
movement in Romania

The Original Fitts' (1951) *MABA-MABA* (*Men Are Best At-Machines Are Best At*) List

- **"Humans appear to surpass present-day machines in respect to the following:**
- 1. Ability to detect a small amount of visual or acoustic energy
- 2. Ability to perceive patterns of light or sound
- 3. Ability to improvise and use flexible procedures
- 4. Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time
- 5. Ability to reason inductively
- 6. Ability to exercise judgment"
-
- **"Present-day machines appear to surpass humans in respect to the following:**
- 1. Ability to respond quickly to control signals and to apply great force smoothly and precisely
- 2. Ability to perform repetitive, routine tasks
- 3. Ability to store information briefly and then to erase it completely
- 4. Ability to reason deductively, including computational ability
- 5. Ability to handle highly complex operations, i.e. to do many different things "at once."
-

MABA-MABA Original Illustration (Fitts 1951)



The Original Fitts' (1951) MABA-MABA (Men Are Best At-Machines Are Best At) List

- “Humans appear to surpass present-day machines in respect to the following:
- 1. Ability to detect a small amount of visual or acoustic energy
- 2. Ability to perceive patterns of light or sound
- 3. Ability to improvise and use flexible procedures
- 4. Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time
- 5. Ability to reason inductively
- 6. Ability to exercise judgment”
-
- “Present-day machines appear to surpass humans in respect to the following:
- 1. Ability to respond quickly to control signals and to apply great force smoothly and precisely
- 2. Ability to perform repetitive, routine tasks
- 3. Ability to store information briefly and then to erase it completely
- 4. Ability to reason deductively, including computational ability
- 5. Ability to handle highly complex operations, i.e. to do many different things “at once.
-

- At present
 - Data are accumulated from various sources
 - sensors
 - transactions
 - web searches,
 - human communication+ Crowdsourcing
 - Internet of Things
 - They contain information useful for
 - Finance, trade, manufacturing, education..
 - The amounts of data are too big for people's storing capacities

Automation[at Large]

Definition: We speak about automation when a computer or another device executes certain functions that the man agent would normally perform.

Remark: Automation has pervaded not only in most safety or time critical systems, such as aviation, power plants or refineries, but also in transportation, libraries, robotized homes, and even intelligent cloths.

Levels of Automation

(Sheridan & Verplanck 1978)

Level 1: The computer

offers no assistance; human must take all decisions and actions

Level 2: The computer

offers a complete set of decision/actuation alternatives, or

Level 3: narrows the selection down to a few, or

Level 4: suggests one alternative, and

Level 5: executes that suggestion if the human approves, or

Level 6: allows the human a restricted veto time before automatic executions, or

Level 7: executes automatically, then necessarily informs the human, and

Level 8: informs the human only if asked, or

Level 9: informs the human only if the computer decides to,

Level 10: The computer decides everything, acts autonomously, ignores the human

Levels for Decision and Action Selection

(adapted from Save and Feuerberg, 2012)

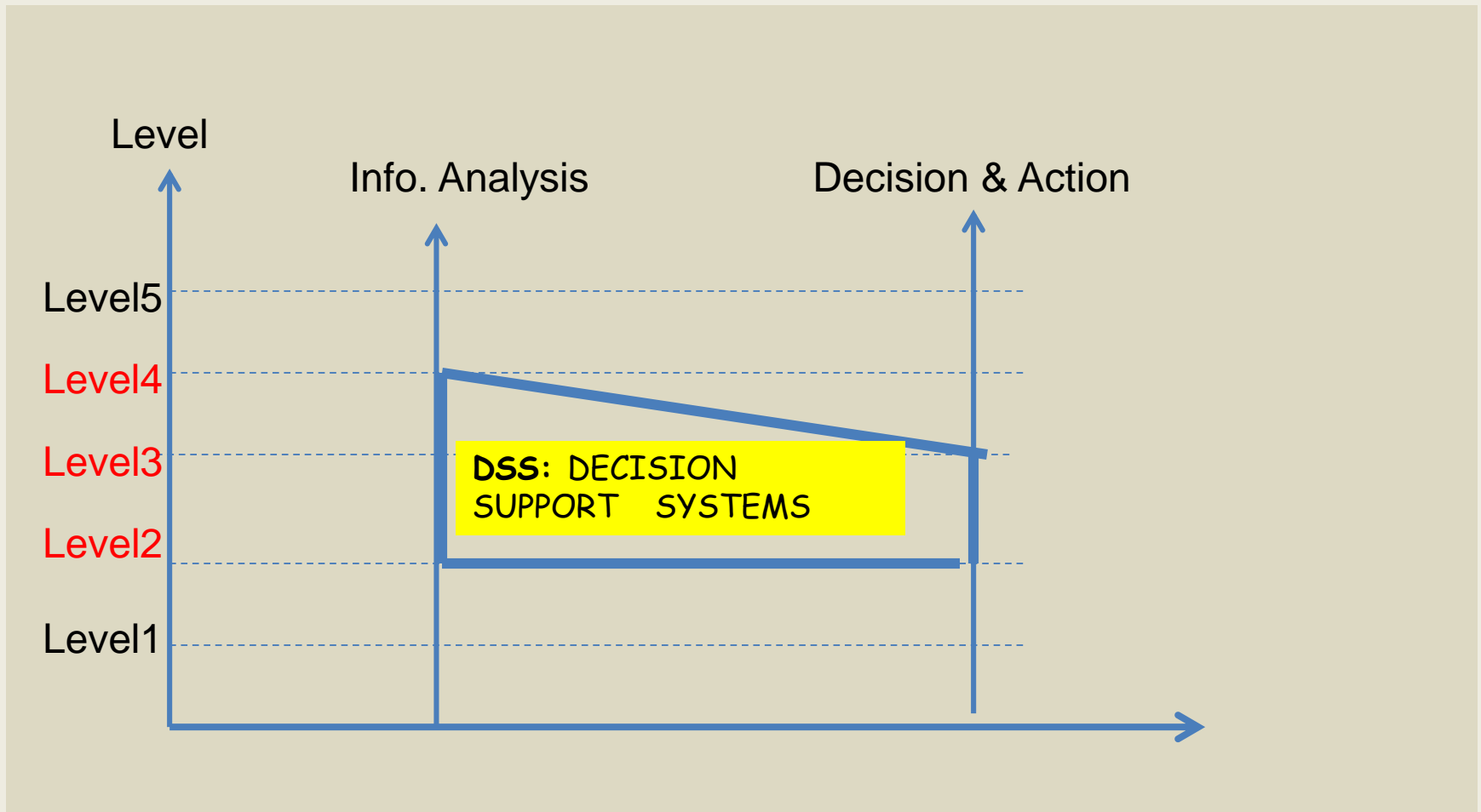
- **L0.** *Human Decision Making: A* (The human agent generates decision alternatives, selects the appropriate ones, chooses the one to be implemented);
- **L1.** *Artefact-Supported Decision Making: A and B* (the human agent decides all actions to implement the selected decision by utilizing paper or other non-digital artifacts (for example, telephone));
- **L2.** *Automated Decision Support: C* (The human agent selects a solution from the set composed of alternatives generated by computer and him/herself);
- **L3.** *Rigid Automated Decision Support: D* (The human agent can select a solution from the set of alternatives generated by the system or asks for new alternatives).

Levels for Information Analysis

(adapted from Save and Feuerberg 2012)

- **L0. Working Memory-Based Information Analysis: E** (The human agent compares, combines and analyses different information items) **and F** (No any other tool or support external to his /her working memory is used);
- **L1. Artefact-Supported Information Analysis: E & G** (Paper or other non-digital artifacts are utilized);
- **L2. Low-Level Automation Support of Information Analysis: H** (Based on user's request, the system helps the human agent to compare, combine and analyze different information items);
- **L3. Medium-Level Automation Support of Information Analysis: H & I** (The system triggers alerts, if the analysis produces results which require user's attention);
- **L4. High-level automation support of information analysis: J** (The system helps the user to compare, combine and analyze different data items concerning the controlled/managed object by using the parameters specified by the user) **and I**;
- **L5. Full automation support of information analysis: I & K** (The system performs comparisons and analyses data available about the controlled object based on parameters specified by the designer or a higher level decision maker) .

Levels relevant for DSS (Decision Support System)



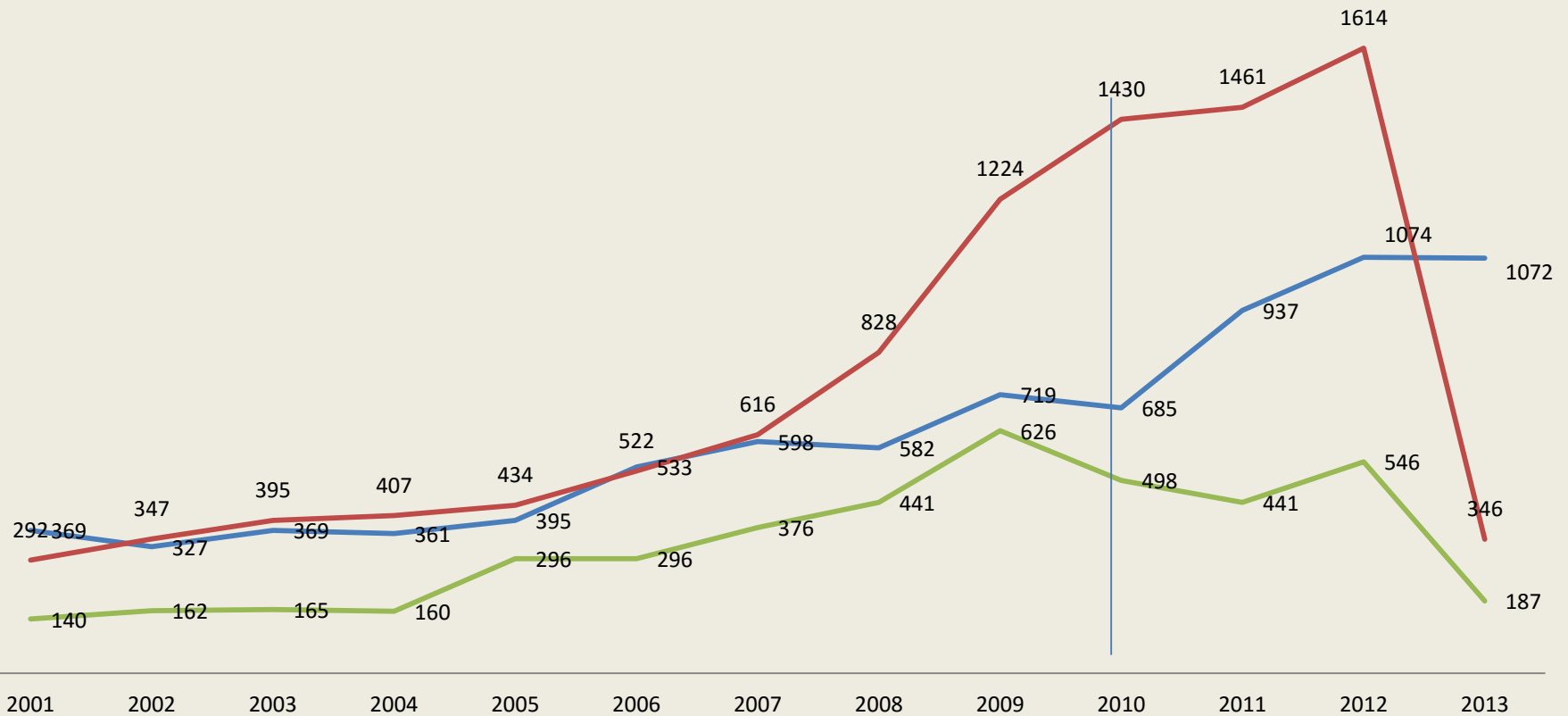
DSS (Decision Support System): A Definition(Filip 2008)

- An anthropocentric and evolving information system which is meant to implement the functions of a human support system that would otherwise be necessary to help the decision-maker to overcome his/her limits and constraints he/she may encounter when trying to solve complex and complicated decision problems that count

DSS in Literature (Filip et al 2014)

Numbers of DSS published materials between 2001-2013

— Science Direct — IEEE Xplore Digital Library — ACM Digital Library



Data-driven DSS Features (Power 2008)

- *Ad-hoc data filtering and retrieval*: drilling-down, changing the aggregation level from the most summarized data one to more detailed ones;
- *Creating data displays* allowing the user to choose the desired format (Scatter diagrams, bar and pie charts and so on) or/and to perform various effects, such as animation, playing back historical data and so on;
- *Data management*;
- *Data summarization*: possibility to customize the data aggregation format, perform the desired computations, examine the data from various perspectives;
- *Spreadsheet integration*;
- *Metadata creation and retrieval*;
- *Report designing, generation and storing* in order to be used or distributed via electronic documents or posted on webpages;
- *Statistical analysis* including data mining for discovering useful relationships.

Major Stages in the Increase of Stored Data Volumes & Associated Technologies

(Hu et al 2014)

- From *Mega* (10^6) to *Gigabyte* (10^9) : in the early 1980's associated with **database machines**;
- From *Giga* to *Terrabyte* (10^{12}): in the late 1980s, associated with the advent of the **parallel data base technology**;
- From *Terra* to *Pentabyte* (10^{15}): in the late 1990s, & **Google file system and Map Redundancy**
- From *Penta* to *Exabyte* (10^{18}) increase of the late years.

Big Data Attributes

(adapted from Kaisler *et al* 2013)

- *Volume* measures the amount of data available and accessible to the organization.
- *Velocity* is a measure of the speed of data creation, streaming and aggregation.
- *Variety* measures the richness of data representation: numeric, textual, audio, video, structured, unstructured and so on.
- *Value* is a measure of usefulness and usability in decision making.
- *Complexity* measures the degree of interconnectedness, interdependence in data structures and sensitivity of the whole to local changes.
- *Veracity* measures the confidence in the accuracy of the data.

Big Data Views

- IBM Infographics: <https://www-01.ibm.com/software/data/bigdata/>
- ORANGE view of Big Data: http://www.eurocloud.ro/wp-content/uploads/bigdata_presentation_Orange.pdf

Business Intelligence (BI) as A Software Platform (Chen et al 2012)

Three classes of functionalities:

- *Integration* : BI infrastructure, metadata management, development tools and enabling collaboration;
- *Information delivery*: reporting, dashboards, ad-hoc query, Microsoft Office integration, search-based BI, and mobile BI;
- *Analysis*: OLAP (OnLine Analytical Processing), interactive visualization, predictive modeling, data mining and scorecards.

BI&A Generations (I)

(Chen 2012)

BI&A 1.0

- adopted by industry in the 1990s,
- predominance of structured data collected by existing legacy systems and stored and processed by RDBM (Relational Data Base Management Systems);
- the majority of analytical techniques use well established statistical methods and data mining tools developed in the 1980s
- The ETL (Extract, Transformation and Load) of data warehouses, OLAP (On Line Analytical Processing) and simple reporting tools are common aspects.

BI&A Generations (II)

BI&A 2.0

- triggered by advances in Internet and Web technologies, in particular text mining and web search engines;
- main technologies: text and web mining techniques associated with social networks, Web 2.0 technology,
- crowdsourcing business practice allow making better decisions concerning both product and service offered by companies and recommended applications for the potential customers

BI&A Generations (III)

BI&A 3.0

- characterized by the large-scale usage of mobile devices and applications such as iPhone and iPad
- the effective data collection enabled by the *Internet of Thing*

Analytic and Data Science Firms

(Gartner's 2015, "Magic Quadrant for Advanced Analytics Platforms")

- *Leaders: SAS, IBM, KNIME, Rapid Miner, ORACLE*
- *Challengers: Dell, SAP;*
- *Visionaries: Alterix, Microsoft, Alpine Data Labs;*
- *Niche players: FICO, Angoss, Predixion, Revolution Analysis, Prognoz, Salford Systems, Tibco Software.*

Five Major Big Data Projects

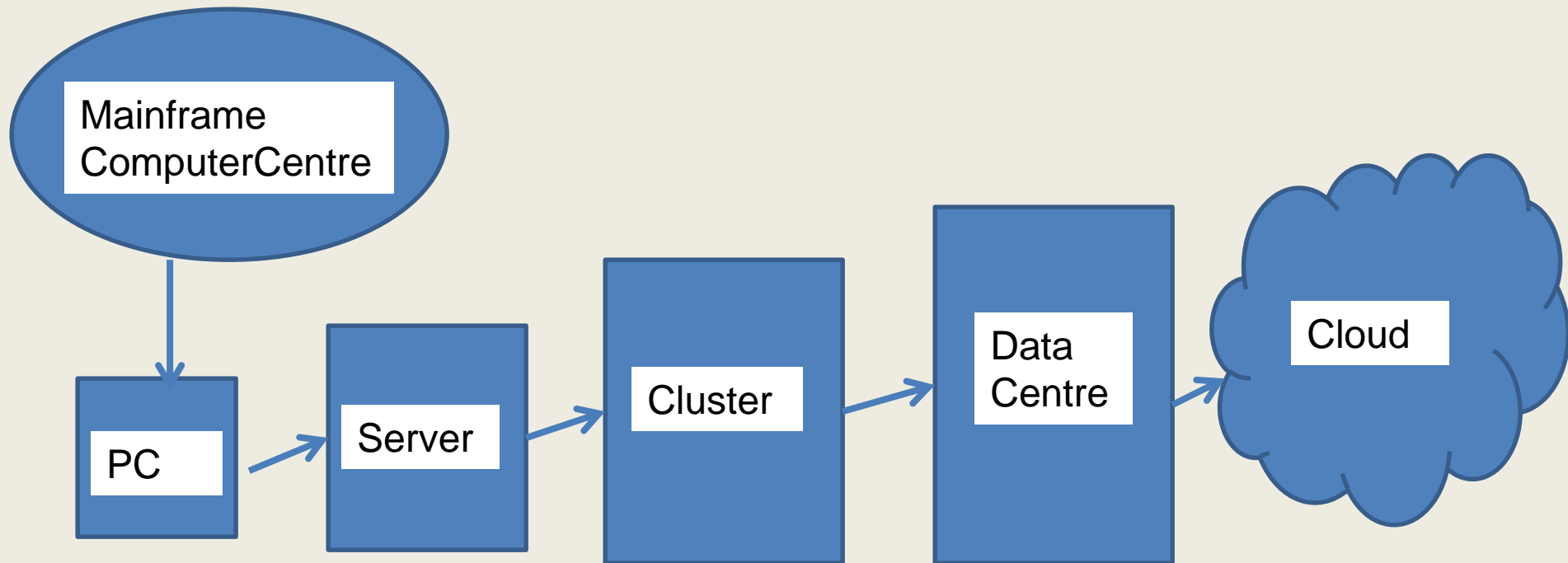
<http://mashable.com/2013/07/17/big-data-projects/>

- Transit NYC (<http://project.wnyc.org/transit-time/#40.75282,-73.95983,14,1011>)
- Topography of Tweets (<https://blog.twitter.com/2013/topography-tweets>)
- Homicide Watch of DC (<http://mashable.com/2013/07/17/big-data-projects/>)
- Falling Fruits (<http://fallingfruit.org/>)
- AIDS vu (<http://mashable.com/2013/07/17/big-data-projects/>)

Cloud Computing: the NIST view

- **Definition:** "CC is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction "(Mell & Grance 2011)
- **Service Models**
 - *Infrastructure as a Service (IaaS).*
 - *Platform as a Service (PaaS).*
 - *Software as a Service (SaaS)*

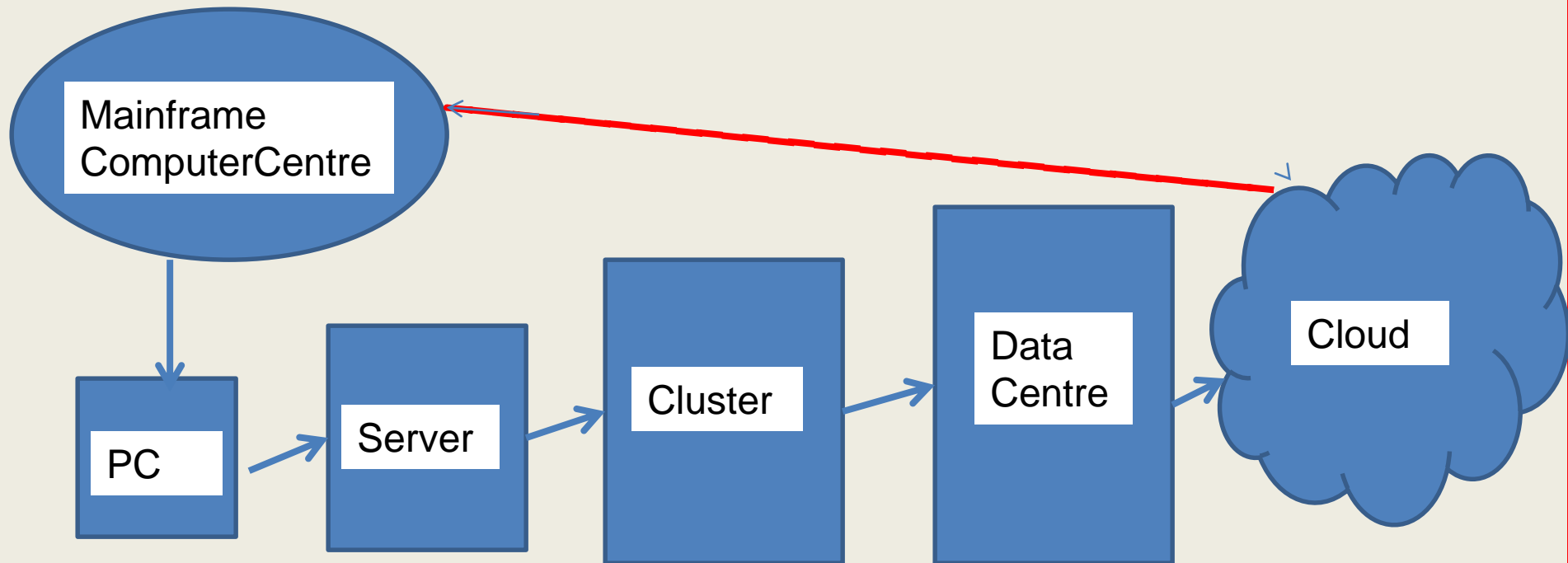
From Mainframe Computer Centre to Cloud Computing



A Remark

- *After all, cloud computing is just mainframe computing in a different way, which is how I learned how to compute when I was a young boy”.*
- From “The impact of disruptive technology: A conversation with Google executive chairman Eric Schmidt “(McKinsey&Company ,2013)

From Mainframe Computer Centre to Cloud Computing



Big Data in Romania(I)

- *A community* of BIG DATA is formed :
<http://bigdata.ro>
- Big Data Events: <https://bigdata.ro/big-data-events/>
- Big Data Week (2013)
 - <http://www.eurocloud.ro/big-data-week/#.UZjIcj4mH3Q>

Big Data in Romania(II)

- *Big international companies* are active and promote their products and services
 - ORACLE: EXALITYCS and ADVANCED ANALYTICS
 - <https://www.oracle.com/big-data/index.html>
 - IBM: Big Data Platform: Apache Hadoop, Stream computing:
 - <https://www-01.ibm.com/software/data/bigdata/>
 - Microsoft : Hadoop

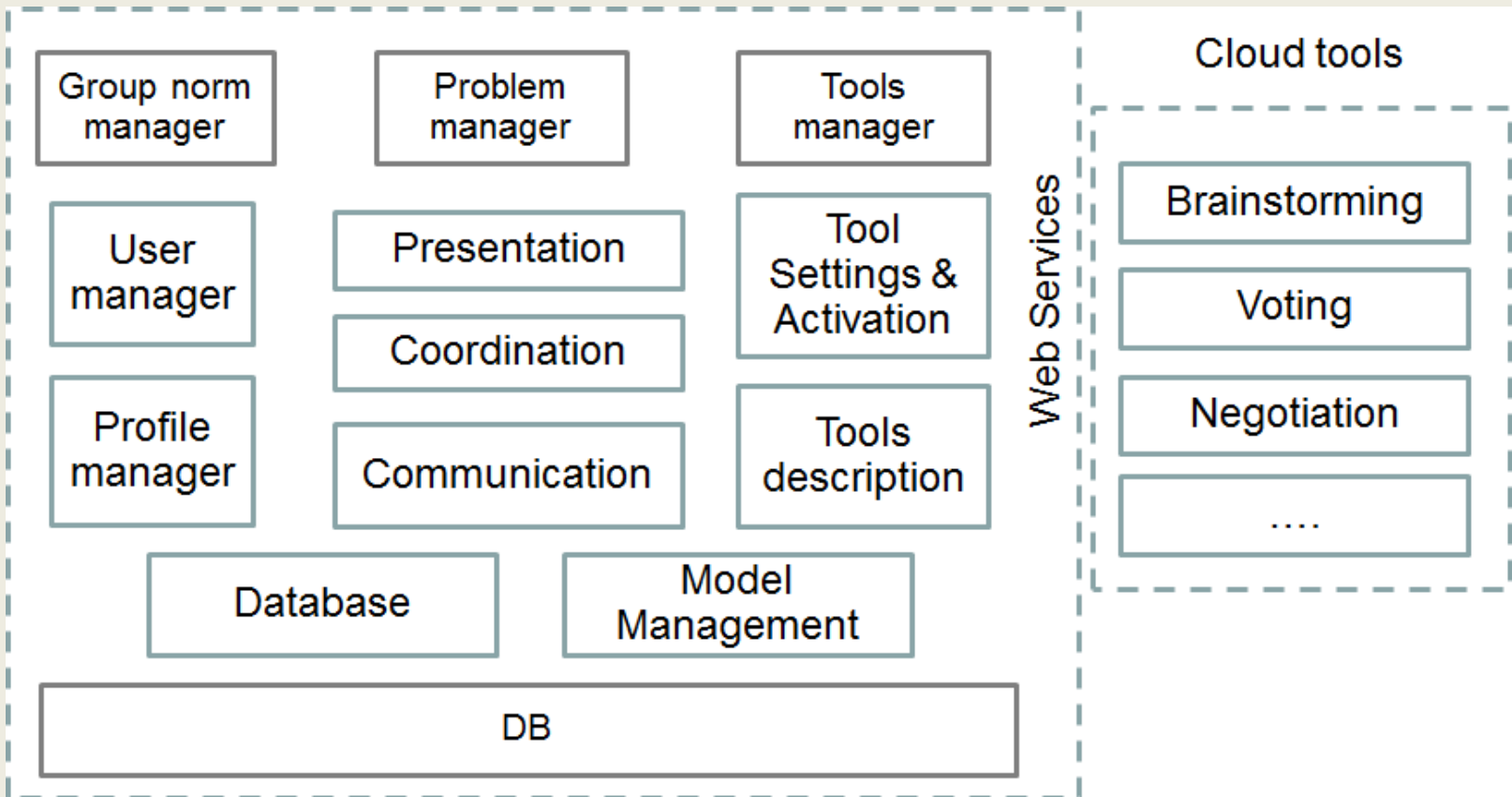
Big Data in Romania(II)

- Telecom Companies : Orange
 - 1,000Exabytes –traffic
 - Finding new business partners
- Intrarom
 - BigStreamer and Next Gen NOC apps (traffic optimization), ValueStreams ...
- The collections of the Romanian Academy Library-BAR (www.biblacad.ro)
- Journalism based on big data analysis
 - Macroscop

BIG DATA and the CLOUD

- Big Data as A Service (BDaaS)
- <http://www.clubitc.ro/2016/02/01/big-data-ca-serviciu/>
- Remark : At present BDaaS
 - For SME (Small and Medium Enterprises) which cannot afford
 - a private cloud
 - A own in house BIG DATA solution,
Only in a private cloud environment of a service provider
 - Barriers : Various billing schemes, bandwidth

iDSS an integrated platform for group Decision Support: the Server(Candea,Filip ,2016)



Labour Market-DSS

(Brandas, Panzaru, Filip, 2016to appear)

- *Import.io* (<https://www.import.io/>) to allow the extraction and conversion of semi-structured data into structured data. The collected data can be exported as CSV (Comma-Separated Values), Excel, Google Sheets or JSON (JavaScript Object Notation).
- *Waikato Environment for Knowledge Analysis (WEKA)* is a machine learning software for data mining processes. It contains tools for data pre-processing, classification, regression, association rules, clustering and visualization (<http://www.cs.waikato.ac.nz/ml/weka/>).
- *Google Fusion Tables* (<https://support.google.com/fusiontables/answer/2571232?hl=en>) : an an experimental data visualization web application that allows the gathering, visualization, and sharing of data tables using Google Maps

General Conclusions

- Data volumes increase dramatically
- BI&A make more effective the *Intelligence* phase of the decision-making process model
- Mobile computing makes possible locating and calling the best experts to perform the *evaluation* of alternatives
- Cloud computing enables *complex computation* during the *Choice* phase
- Social networks enable *crowd problem solving*
- *The combination of Cloud computing and BI&A is the solution for Big Data problems*

References (I)

- Chen H, Chiang R H L, Storey V C. (2012) Business Intelligence and Analytics: from Big Data to Big Impact. *MIS Quarterly*, 36(4), December: 1-24
- Candea, C, Filip F G (2016) Towards intelligent collaborative decision support platforms. *Studies in Informatics and Control*, 25(2): 143-152
- Dekker SW, Woods DD (2002) MABA-MABA or abracadabra? Progress on human–automation co-ordination. *Cognition, Technology & Work*, 4(4), pp.240-244.
- Filip FG, Suduc AM, Bîzoi M. (2014) DSS in numbers. *Technological and Economic Development of Economy*, 20(1): 154-164.
- Fitts, P. M. (1951) *Human engineering for an effective air navigation and traffic control system*. Washington, DC: National Research Council.
- Hu H, Wen Y, Chua T-S, Li X (2014) Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. *IEEE Access*: 652-687..
- Kaisler S, Armour F, Espinosa J. A, Money W (2013) Big Data: Issues and Challenges Moving Forward. In: *46th Hawaii International Conference in System Sciences*, IEEE Computer Society: p. 995-104
-

References (II)

- Mell, P., Grance, T. (2011). *The NIST definition of Cloud Computing*. Special publication 800-145 (<http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>)
- Rasmussen J (1983) Skills, roles and knowledge, signal, signs and symbols and other distinctions in human performance model. *IEEE Trans. Systems, Man and Cybernetics, SMC*, 13(3): 257-266
- Sajithra K, 2Dr. Rajindra Patil (2013). Social Media – History and Components, *JOSR-JBM*, 69-74
- Save L, Feuerberg B (2012) Designing human-automation interactions: a new level of automation taxonomy (<http://hfes-europe.org>).
- Sheridan T B, Verplank W (1978) *Human and Computer Control of Undersea Teleoperators*. Man-Machine Systems Laboratory, Dept. of Mechanical Engineering, MIT, Cambridge, MA

*Thank-you for your
attention!*

