DATA AXIS MUNDI

Professeur Serge Miranda (UCA)

Seminar is going to start at 2pm (CET)

DATA AXIS MUNDI

(Strategic vision on (BIG) DATA Economy around technical disruptions)

Professeur Serge Miranda

(miranda.serge@gmail.com)

Director of MBDS (and eMBDS) Master degrees at UCA

www.mbds-fr.org & MS BIHAR (and eBIHAR) at ESTIA

Series of ESTIA webinars on

« AI and BIG DATA »

for students from MBDS, BIHAR, eBIHAR (and its GRADEOs), Miage (ANR THEMES) and …others

To register :

https://forms.office.com/Pages/ResponsePage.aspx?id=vCENNyqVPkmiyRPPoeY1IMbaYPunPfpDvd2_5dohiIhUOEhMVlMzR0tXTkU0RzdIQTBZNlhIMVgxVS4u&qrcode=true

ESTIA Webinars on

Artificial intelligence and Big Data

MSc BIHAR MSc (2020-2021)

Open to graduate students and companies

2pm-4pm (CET) – on TEAMS (and ESTIA3 Amphi , Technopole IZARBEL, 64210 Bidart France)

10/10/2020 « *TENSORFLOW BY EXAMPLE (Tensorflow, Keras and Colab)* » Alison TEMIN (ESTIA et UCA) *< special one-day seminar >*

08/12/2020 → « *DATA AXIS MUNDI (Strategic vision on AI and Big Data)* » **Pr Serge Miranda (UCA)**

17/12/2020 → « *BIG DATA ANALYTICS in SPORTS* » **Pr Alex Rayon (Univ. Deusto, Spain)**

04/01/2021 → 6pm-7.30 pm (CET) : « *Cybersecurity survey* » **Pr Alban Gabillon (Univ. de la Polynésie Française, Tahiti)**

29/01/2021 → « *AI Strategic vision and research* » **Pr Marco Gori (Univ. of Siena, Italie)**

05/02/2021 → « *Past and Present of Deep Learning: Now What?* » **Pr Stefano Melacci (Univ. of Siena, Italie)**

26/02/2021 → « **AI in action in manufacturing 4.0** » (« *L'Intelligence Artificielle en action : cas d'applications issus du monde de l'Industrie, Energie et Transport* ») **Romain Roquefere (HUPI, Technopole Izarbel)**

04/03/2021 → « *Machine learning with a data pipe line for video-streaming analysis aiming at M6 customer segmentation (and Facebook)* » **Dinh Duy Nguyen (BEDROCK Streaming)**

To register :

https://forms.office.com/Pages/ResponsePage.aspx?id=vCENNyqVPkmiyRPPoeY1IMbaYPunPfpDvd2_5dohiIhUOEhMVlMzR0tXTkU0RzdIQTBZNlhIMVgxVS4u&qrcode=true

MBDS and BIHAR : Two CS masters on DATA Technologies with an INNOVATION laboratory on USAGE engineering (disruptive ICT technologies)

ESTIA
INSTITUTE OF TECHNOLOGY

CONFÉRENCE DES GRANDES ÉCOLES

MBDS

BIHAR

MBDS
(Mobiquitous BIG-DATA Systems)

(Big data Intelligence for Human Augmented Reality)*

e-BIHAR

e-MBDS
Changing the world of professional learning

* « *BIHAR* » means « *TOMORROW* » in Basque Language

e-BIHAR

The two first MOOC-based European master degrees in CS in Europe : eMBDS (UCA) in 2019 and eBIHAR (ESTIA) in 2020 with ORACLE as a professional partner and great feedback

The two first GRADEOs (micromasters) in Europe on FUN MOOC plaform on BIG DATA and AI on January the 4th < for continuous education>

« I'm writing to you to express my gratitude for the courses I took at University of Nice. It was great to be a part of E-mbds program. I was very pleased with the selection of courses as they covered the latest and greatest technologies on the market. I also was impressed with professors' knowledge level, and amount of work they put into their classes. I hope I will have an opportunity to take other classes at your University in the future. Please keep me informed on your future programs. »
—Irina Mok, Oracle Senior Manager Software Development (Oracle USA)

« It was awesome being a part of the eMBDS learning program! I never imagined MOOCs and online learning could be so effective and was apprehensive initially when i signed up but as I progressed through the courses, i felt more confident that this degree would definitely help enrich my skills and help me at my work place and further educational endeavors. Some of my personal favorites were the courses on Javascript, Native Mobile Programming and the ones on Data Mining and Machine Learning. In addition, the Oracle University courses were also very detailed with hands-on lab sessions. A big thank you to (Datum Academy) namely Mishket Hamida, Betty Schmitt for being available and patiently answering all my questions on the administrative part and to all the professors for the wonderful MOOCs, live sessions and support on the courses."
—Amrita Panda, Principal QA Analyst (Oracle USA)

"I'm very pleased that I was one of the first students who graduated Master of Computer Science degree program based on MOOCs. The eMBDS program offered by the University of Nice with a partnership with Datum academy and Oracle University will boost your technical knowledge and allows you to master all hands-on expertise in emerging technology. At the same time, you will have a chance to adopt Oracle technology and gain professional certificates from Oracle University. It is big deal. Thank you very much to all of our professors, faculty staff, facilitators, and their support and collaboration. I can recommend this unique opportunity to anyone who wants to advance their career and deepen their IT credentials. "
—Zolbayar Zorigoo, Orale Senior Sales Consultant (Oracle Romania)

"Course content is very apt and adheres to current IT trends, I really liked the graduate course content. It includes Big Data, Mobile Technologies, Javascript, Database, Agile, NFC, Blockchain etc. All the courses included in eMBDS are must for any IT Professional. Really liked the content of each of the courses : it covers all the basic and intermediate levels of content. Later individuals can scale themselves to Expert Level. As a conclusion, eMBDS is a very good on-line master degree for any person who wants to improve their skills and keep themselves updated; it includes all the latest technologies. I would highly recommend taking it up mainly because of the course content and the great support from Datum Academy. This online course would be difficult or impossible without great support from Datum Academy".
—Dharmendra Singh, Oracle Principal Software Engineer (Oracle India)

eMBDS

eBIHAR

DATA AXIS MUNDI !

« AXIS MUNDI ? »

2 questions before starting…

- **Q1 : Quantity of DATA produced since the beginning of humanity (5000 years ago)? Produced evey second on Internet ?**
- *Quantité de données produite depuis le début de l'écriture il y a 5000 ans (Google)? et aujourd 'hui par seconde (UC Berkeley) ?*

- **Q2 : Quantity of data to get a digital twin ?**

Quantité de données utile pour construire un double numérique « complet » (Jim Gray et Digital Immortality) ?

- **1 TERA BYTES : 10**12**
- **1 PETA BYTES : 10**15**
- **1 EXA BYTES : 10**18**
- **1 ZETA BYTES : 10**21**
- **1 YOTTA BYTES : 10**24**

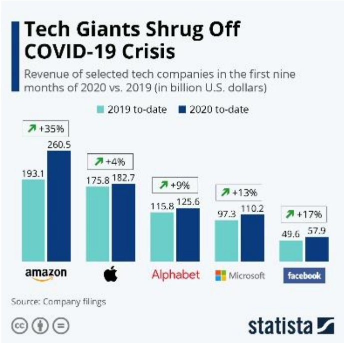
March 2020 and COVID Confinement ➔ paradigm switch in a new data-centrics digital era of ANTHROPOCENE

- Digital transformation: e-services, e-commerce, NFC payment, e-life
  - eHEALTH : Doctolib with 100 000 queries a day (April 2020)
  - Remote working : Ex : Zoom : 300 million meeting connections a day (in April 2020) < 10 million in Dec 2019> ; 1st app on Appstore (August 2020)
  - Blended Learning : ***from centripete university to centrifugal multiversity***
  - *zoom and Moocs!*


« On a découvert le confinement ici »

ESTIA Alors que les élèves ingénieurs ont déserté le campus, Franck et Jean-Charles suivent leur master en distanciel dans leur salle de classe

NFC

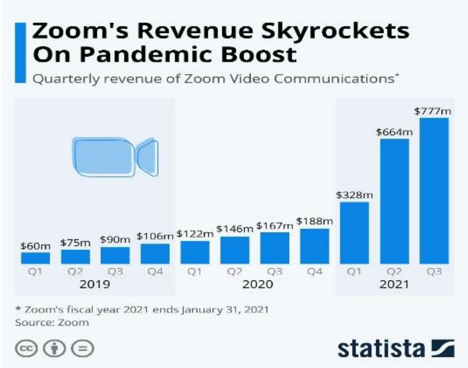



Wall Street jette le masque

AU RYTHME où ça va, Apple vaudra bientôt plus cher que tout le CAC 40 réuni ! A lui seul, il dépasse les 1 500 milliards de dollars. Microsoft itou. Et Amazon vient de passer les 1 000 milliards. Jamais Wall Street n'avait autant pété la forme. Il faut remonter à 1998, remarquent « Les Echos » (1/7), pour trouver un trimestre **« aussi flamboyant »**.

On a failli s'inquiéter. Le Covid a entraîné une chute **« vertigineuse »** des marchés financiers. Mais le rebond est **« spectaculaire »**. En un tournemain, les effets du krach ont été aux trois quarts effacés. Il faut dire que les fameux Gafam, les mastodontes américains du Net et du high-tech, s'en sont mis plein les poches. Le télétravail, le téléenseignement, les téléréunions, la télémédecine, la téléconnerie, le télécommerce ont connu un succès foudroyant.

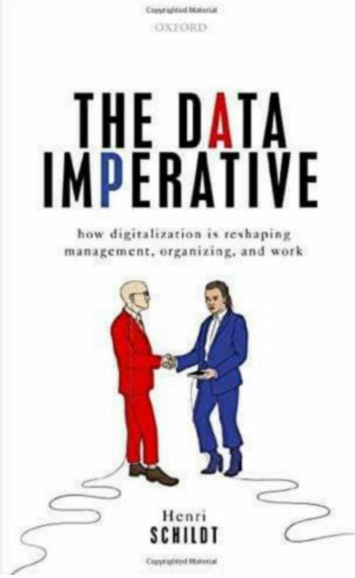
On attend la deuxième vague avec impatience…

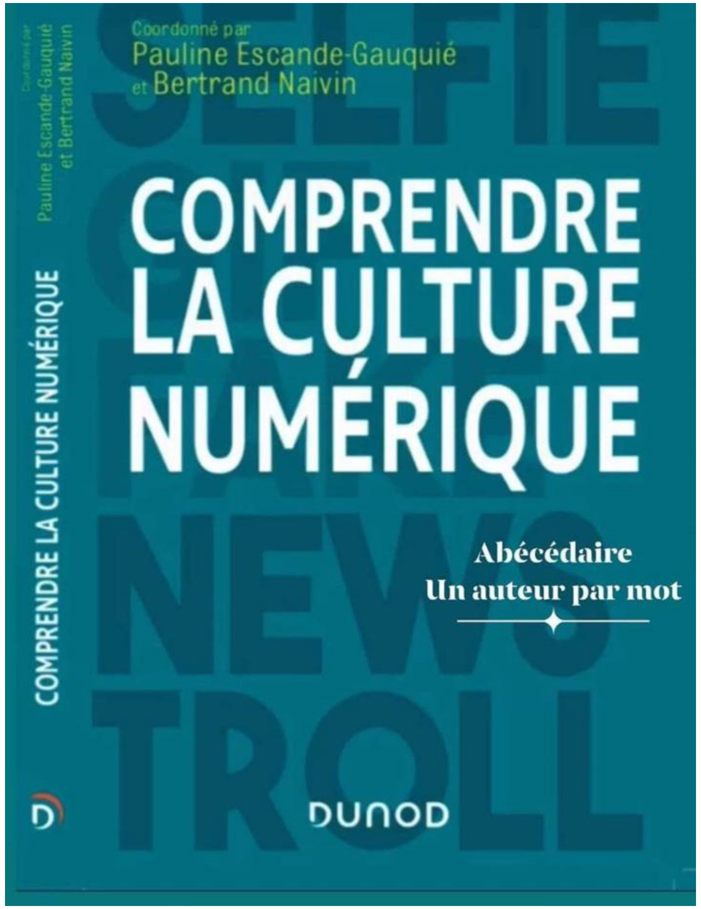


e-BI

Some news in the last six days and WEBINAR CONTENTS

- December the 2nd :
  - **blended learning**
    « ADIEU aux AMPHIS (***Farewell amphiteatres***) » LE MONDE Diplomatique, December the 2nd : « ***May 1968 dream it and virus made it*** »
  - **ARTIFICIAL INTELLIGENCE and medecine** https://siecledigital.fr/2020/12/02/sante-intelligence-artificielle-deepmind-proteines
- December the 4th : Tweet on a QR Code built with 130 000 trees in China !
- December the 5th : New book :
  « ***The DATA imperative*** » on BIG DATA applications
- December the 7th : *SAS Research report : « Innovation : from DATA to BUSINESS »*





PLAN

Disruptive technologies in the DATA economy
- **DATA (IOT, Smart objects) & BIG DATA**
- **<u>AI and DATA learning (ML & DL)</u>**
- ***NFC, LIFI, Blockchain,..etc***

DATA PARADIGMS : SQL & NO SQL
Three types of DATA : structured, semi-structured and *unstructured*

Three dimensions in the DATA economy
- **BOTTUM UP paradigm**
- **little Big Data**
- **smart places (DATA Spaces)**

Conclusion : “Spiralist” innovation in the DATA economy

Part of introduction of the Mooc on “Distributed big data management” (GRADEO in January 2020); eBIHAR MSc on FUN platform

Partie aussi du Chapitre du Livre pluridisciplinaire DUNOD de Sept 2019 & (Keynote, Assises CNAEMO, TOULOUSE, 30 septembre 2020)

RAW RESOURCE of this millennium like love and happiness ?

***English – "Happiness is a marvellous thing : the more you give, the more you are left with"* (Pascal)**

French – « Le bonheur est une chose merveilleuse : plus tu en donnes plus il t'en reste » (PASCAL)

Arabic: كل تيقب املك اهتيطعأ املك ليمج ءيش ةداعسلا

Creole (Haïti) - Ala yon bèl bagay se kontantman, plis ou bay ladan'l plis ourete ladan'l!

Russian : Счастье – волшебная вещь: чем больше ты его даришь, тем больше тебеостаётся»

Spanish - la felicidad es un artículo maravilloso: cuanto más seda, más le queda a uno

Occitan - la felicitat una chausa meravelhosa:mai ne'n donas, mai te'n rèsta

Swedish- lycka är något underbart:ju mer du har att ge, desto mer har du kvar av den.

Italian - la felicità è qualcosa di meraviglioso: più ne dai e più te ne rimane

German- Glück ist eine wunderbareSache: je mehr du schenkst, destomehr hast du

Roman - 'a felicità è quarcosa de meravijoso: più 'a dai e più te ce rimane

Hungarian- a boldogság csodálatosdolog: minél többet adsz belõle, annáltöbb marad neked

Brazilian Portuguese - a felicidade é uma coisa maravilhosa: quantomais você dá, mais você recebe

DATA / INFORMATION

- **Properties ☺ ?**

  1-1 = 2 !

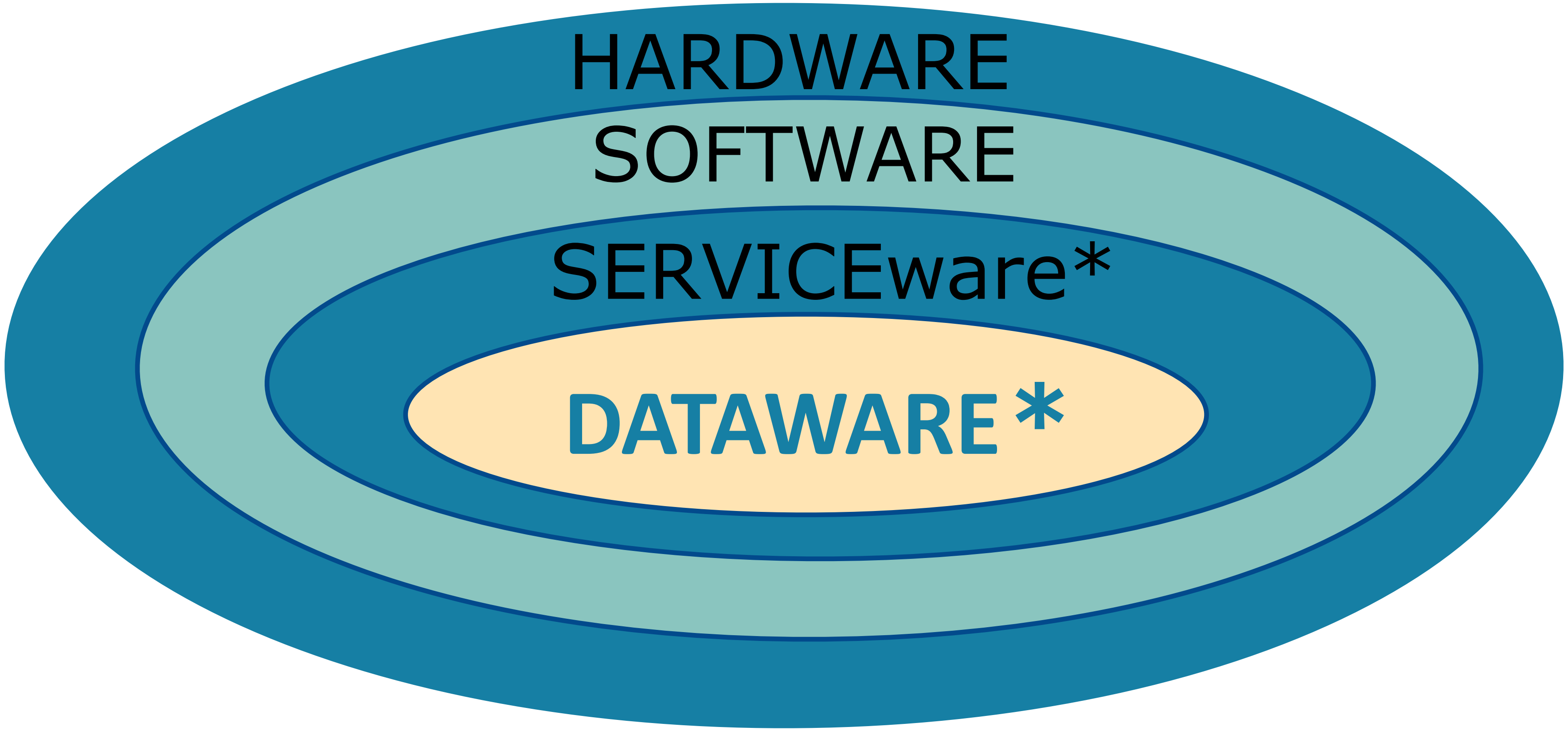
- **1+1 = 4 (METCALFE ‘s law)**

« PAUVERTY is DATA-access denial »

F. Verella (Haïti)

World Bank (Washington)

28 years of innovation at MBDS with 4 Copernician digital revolutions !



* "*USERWARE*" : « BIG DATA & DATA SCIENCE »

Some visions of the future of big-data management

INSTITUTE OF TECHNOLOGY

- **CLOUD COMPUTING**
  - INFRASTRUCTURE as a SERVICE (IaaS)
  - PLATFORM as a SERVICE (Paas)
  - **DATA as a Service (DaaS)**
    from Oracle ;
    ANALYTICS as a SERVICE (AaaS)
    from Google, IBM, etc.

- ***« CAMS »*** *(IBM 2014)*
  - **C**LOUD for servers
  - **D**aaS/**A**aaS :
    « (DATA) ANALYTICS as a service »
  - **M**obility (smartphones applications)
  - **S**ocial Networks (for data integration)

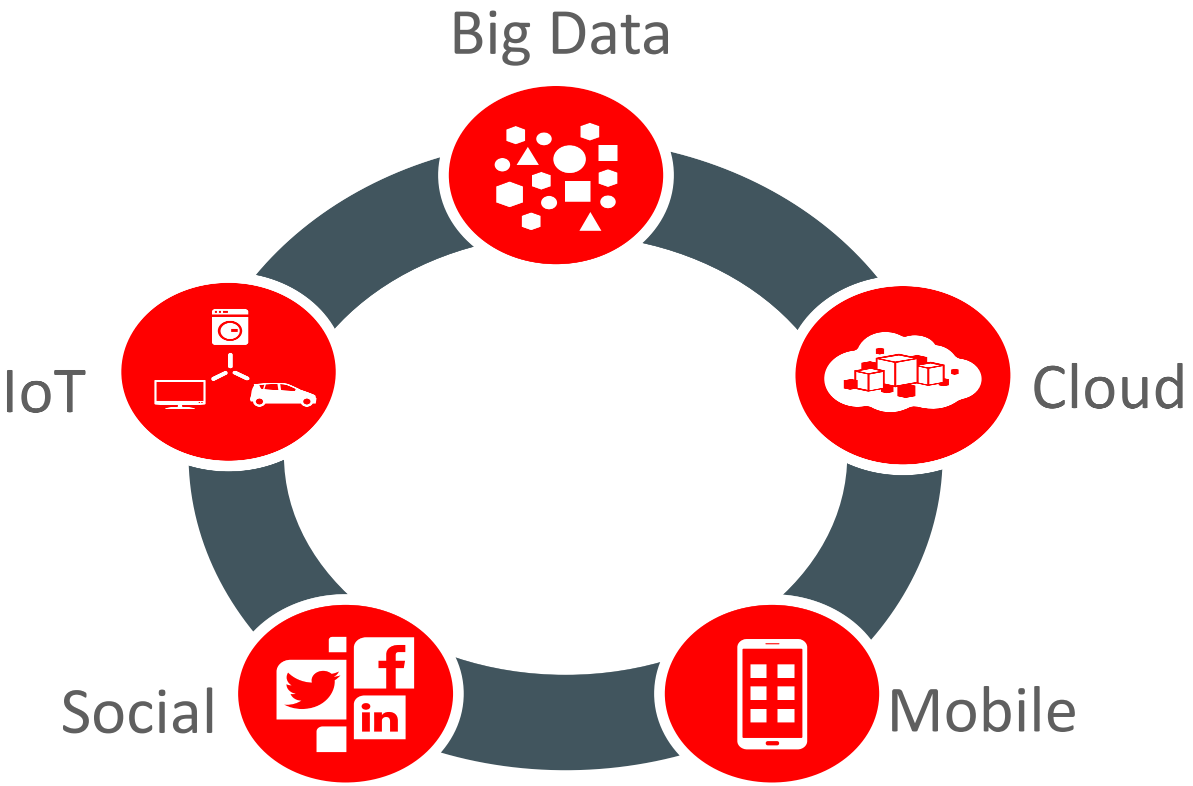
- ***« SMAC »*** *stacks (CITY GROUP, Vikram Pandit)*

« No business model in the future could succeed without the DATA »

Social
Mobile
Applications
Cloud

e-BIHAR

ORACLE vision : CI-MBDS☺



A strategic vision of the digital DATA revolution

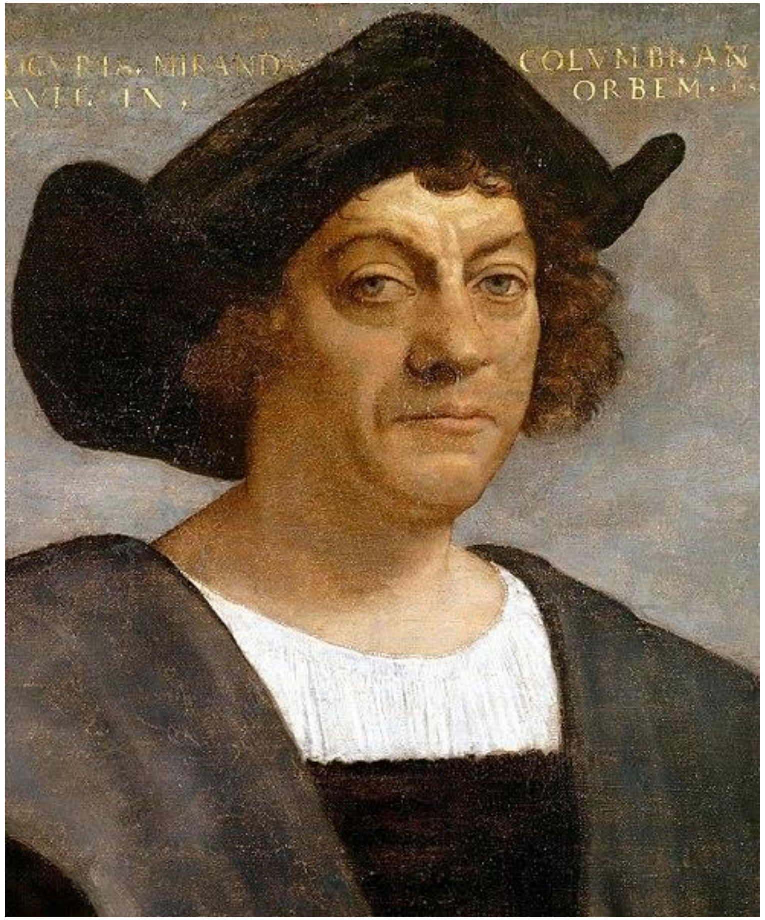
ESTIA INSTITUTE OF TECHNOLOGY

PASSION hexagram
(in RAKU by Marina Latta for MBDS 20th anniversary)

e-BIHAR

To be AUDACIOUS : « Be OUT of the box » !

INSTITUTE OF TECHNOLOGY

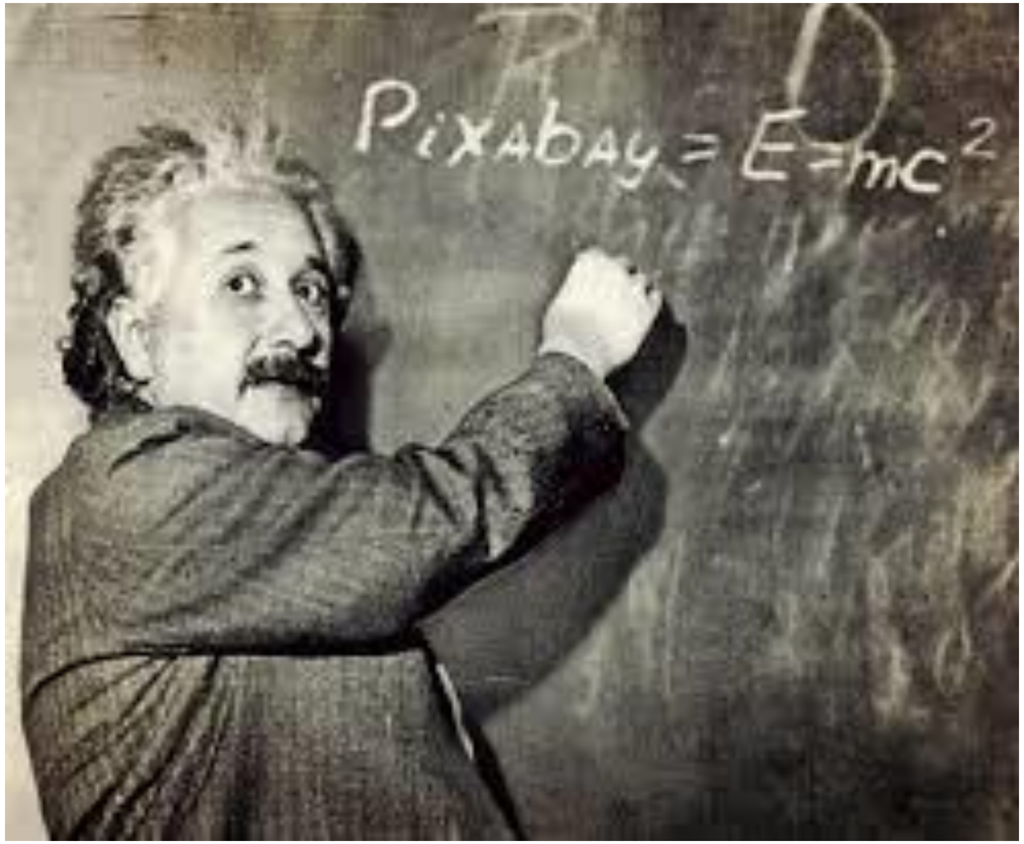


- « *Your Highness …determined to send me, to the country of India…and furthermore directed that I should not proceed by land to the East as is customary, but by a Westerly route,* **in which direction we have hitherto no certain evidence that anyone has gone** »

- Christopher COLUMBUS

e-BIHAR

Michel Serres on 5 key contributors to humanity and …



- Moses : « ***Everything is LAW*** »
- Jesus : « ***Everything is LOVE*** »
- Marx : « ***Everything is MONEY*** »
- Freud : « ***Everything is SEX*** »
- Einstein : « ***Everything is RELATIVE*** »

&…

- Jim GRAY:

« *Everything is DATA* »

« DATA » as prefix or suffix !

INSTITUTE OF TECHNOLOGY

DATA as prefix:

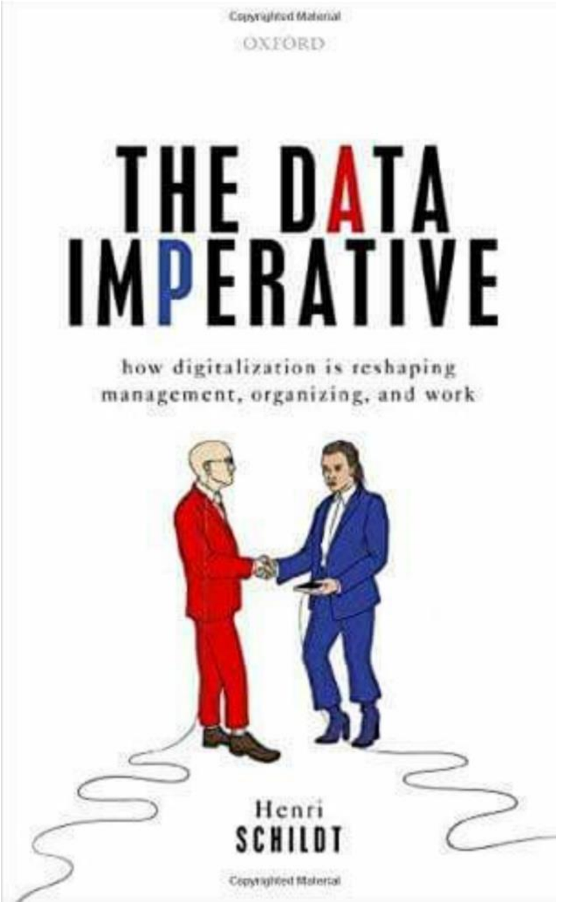
DATA base (19/8/1968 : Ted Codd and relational data model), DBMS
DATA Schema
DATA bank, DATA STORE
DATA warehouse
DATA mart
DATA mining (OLAP, Correlations, ..), DATA Analytics, DATA Pumping (ETL), Little Big Data
DATA Systems
DATA mash up
DATA SCIENCE
DATA LAKE
DATA WEB

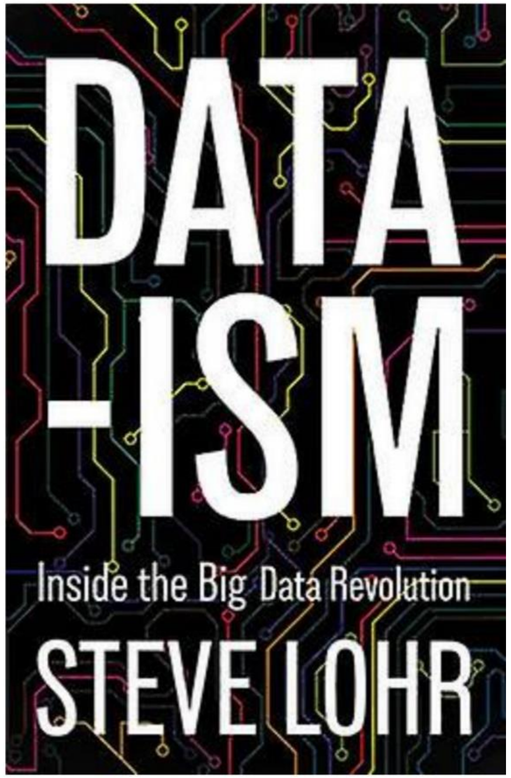
DATA as suffix:

Relational DATA
Linked DATA,
Labelled DATA
Meta DATA
Open DATA
Web data
Smart DATA
BIG DATA and

new DATA-centrics jobs
Big Data architect
CDO « Chief DATA Officer »,
« DATA SCIENTIST »,
« DATA BROKER »
« **DATA ENGINEER** »

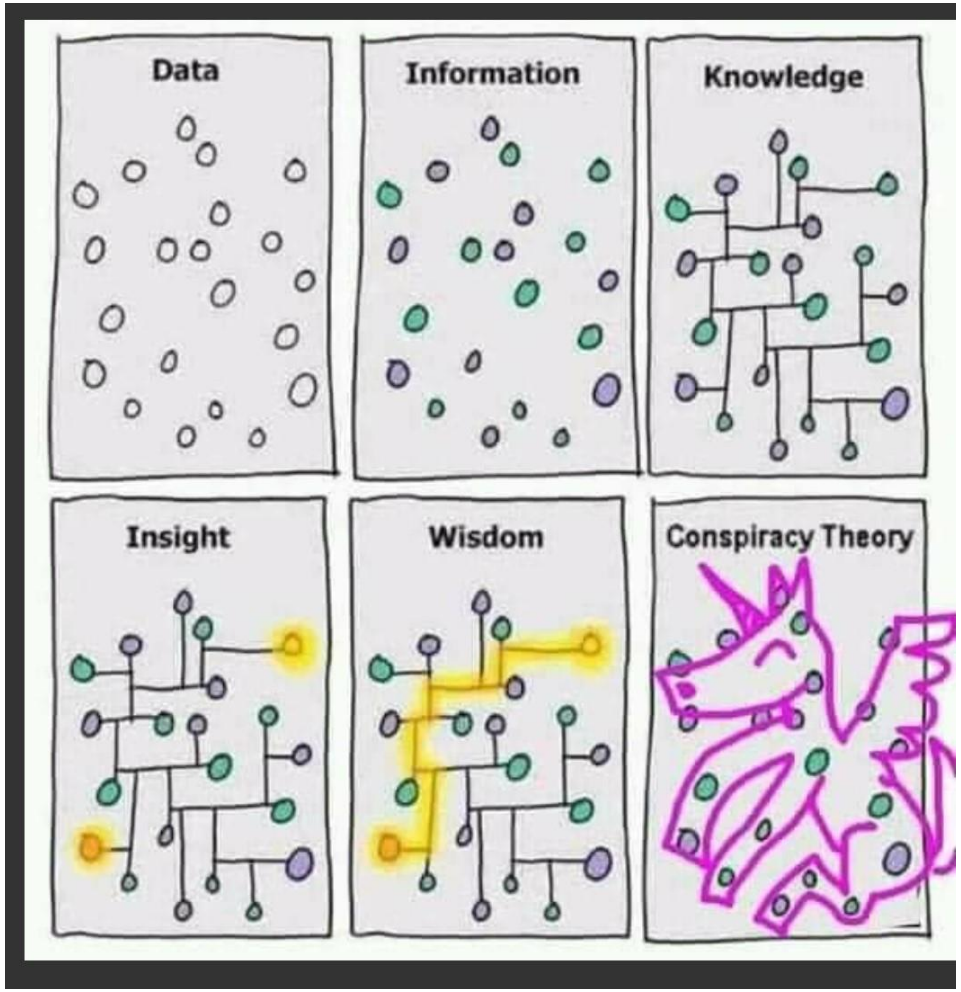
Data-centrics Economy & DATA « deluge »





DATA :
next BIG WAVE

INSTITUTE OF TECHNOLOGY



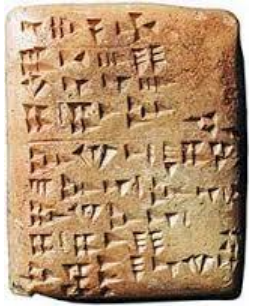
DATA & two major human defaults !

1. **Man forgets**

2. **Man can be wrong**
   « *Errare humanum es* »

DATA ?

- ***Recording of ANYTHING (fact, measure, video,..) on ANY SUPPORT with a given code***




Spirals on La Piedra Escrita Jayuya Puerto Rico

DATA ?

« Value of man consists in GIVING » Einstein
« It is DATA i.e. STUPID » Jim Gray

- « **DATUM** » ➔ « **DATA** » from DARE ***(to GIVE in Latin language)*** *« what is given »*

- **symmetric word : « CAPTUM » ➔ INFORMATION** !

➔ « **CAPTA** » (Kitchin 2014) from Captio (to TAKE in Latin)

« DATA »→ « INFORMATION » (*CAPTA*) → KNOWLEDGE (*Insight*)

KNOWLEDGE ?

3 Types of Knowledge for ARISTOTELES :

1. **EPISTEME**
   SCIENCE **<Knowing>**
2. **TECHNE**
   TECHNOLOGY **<Knowing to do>**
3. **PHRONESIS**
   ETHICS
   <Knowing to be ; wisdom>

« SMART OBJECT » =
DATA + COMMUNICATION

(SMART City, car, home, bus stop, museum...)

- **SMART OBJECT :**
  DATA (Production/Processing/ retrieval) + **COMMUNICATION**

- **LIFE IN BIOLOGY ?**
  LIFE = **INFORMATION + COMMUNICATION**

Smart objects and IOT *(Internet of Things)*

- 50 billions of smart objects by 2020 (Gartner)
  - from IOT to IOE (Internet of EVERY THING) :
  - every object will become smart !

- IoT is a catalyst for Big Data

BIG DATA ?

11 BIG DATA facts SHAPING the future ! (www.graziti.com)

INSTITUTE OF TECHNOLOGY

- #1 :90% of DATA in the world were created in the last 2 years
  - By 2020 : **1.27 MB of data created EVERY SECOND for every human being**
- #2 : By 2020 : Global IP traffic will reach 44 ZETA BYTES
- #3: By 2020: **30% of web browsing will be done WITHOUT A SCREEN**
- #4: By 2020 100 million customers will shop thru augmented reality (and chatbots)
- #5: in 2018, 3 million workers supervised by a ROBO-BOSS
- #6: **By 2020 algorithms will alter the job of 1 billion workers**
- #7 : By 2020, 40% of employees will cut their healthcare costs by weering health trackers (smart watches)
- #8 : By 2022, IOT will get consumers and business 1 trillion dollars a year
- #9 : By 2019, 20% of brands will abandon their mobile applications
- #10 : By 2020, 40 million cars will use ANDROID
- #11: **By 2020, 85% of Internet transactions will be processed by machines** (GARTNER)»
  - Forecast of 75% grow on AI/CHATBOT from 2016 to 2021 (OVUM)

e-BIHAR

« The 3 V » of BIG DATA ?

ESTIA INSTITUTE OF TECHNOLOGY

DATA deluge, DATA Tsunami, DATAnome, Dataware, DATAISM,…

- **The 3 « V »** (M.Stonebraker)
  - **Volume,**
  - **Variety**
  - **Velocity**

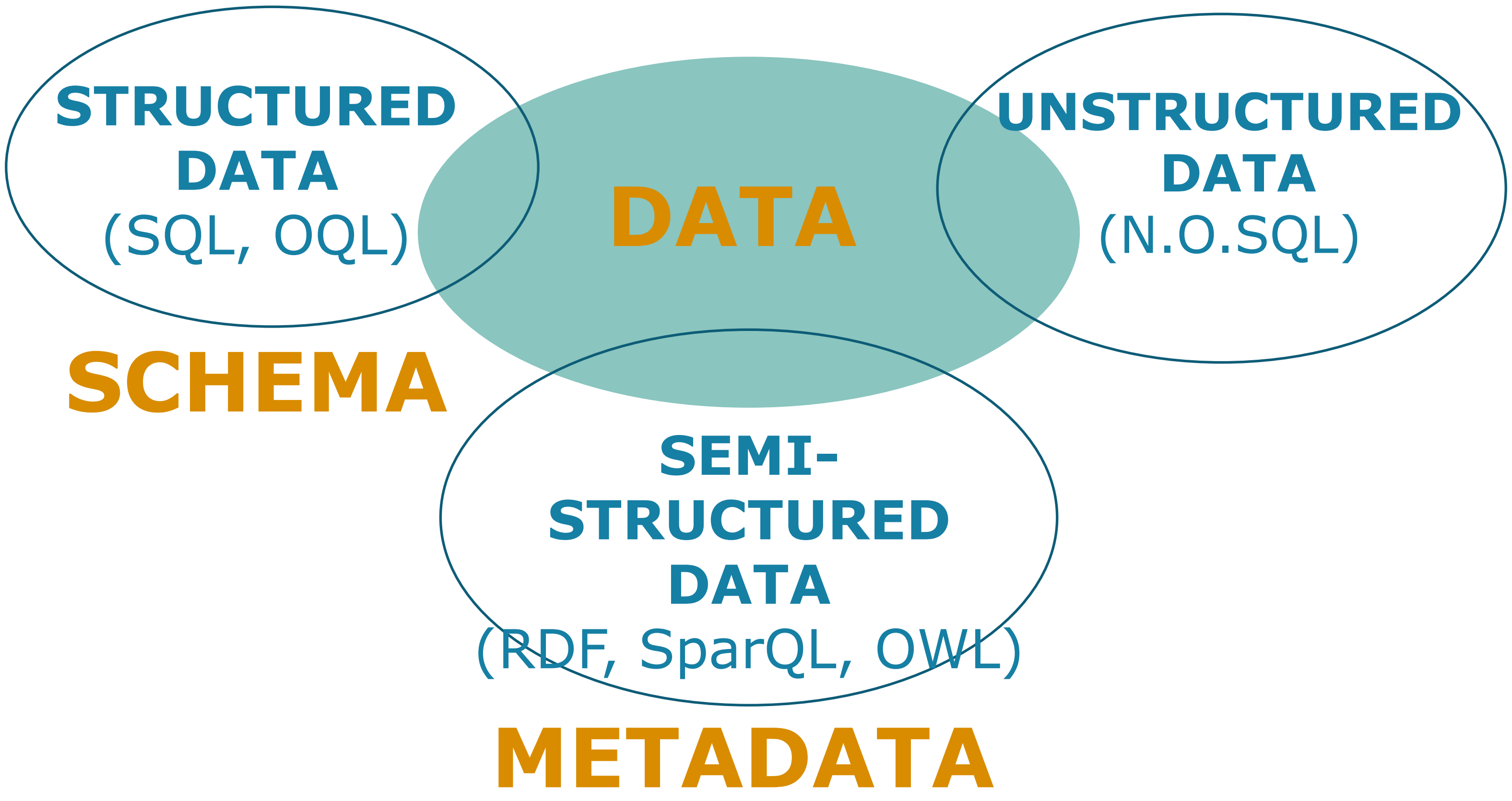
e-BIHAR

« VOLUME » (« BIG » data) : ➔Exa-bytes per second !

- **5 exa*bytes of data** have been produced since the beginning of humanity until 2010
  - ***« Today we produce that amount every 2 days »***
    Eric Schmidt
    (Google CEO, Davos 2010)
  - in 2013 : every 10 minutes !
  - **In 2015 : every 10 seconds (UC BERKELEY report, 2013)**
  - In 2020… ZETA* bytes..

*** NOTE :**
PETA : 1015; EXA : 10**18; ZETA : 10**21**

« VARIETY » (BIG DATA) : 3 types of DATA in CS



« VELOCITY » (Big Data) : DATA flowing

- DATA from IOT (Captors, sensors)

- Every second on INTERNET : https://www.internetlivestats.com/one-second/ (2020) http://onesecond.designly.com/, http://www.internetlivestats.com/
  - 9200 tweets (400 M a day)
  - 1000 pictures loeaded on Instagram
  - 100 terabytes of Internet traffic
  - 90 000 videos seen on youtube (> *1 hour of video posted on YOUTUBE every second)*
  - 90 000 queries on Google <Google reported earning more than $50 billion in ad revenue in 2013, netting them **around $1,602 in profit per second**
  - 3 M de mails
  - 5000 skype calls
  - 10 000 meetings on ZOOM (300 M meeting a day on Zoom during the COVID; 10 Millions 3 months before)
  - *3000 visits to Facebook (>3 billion friends end of 2020)* < Every day, Facebook users like an average of 4.5 billion posts, share more than 4.7 billion status updates with their friends or followers, and watch over 1 billion videos.>

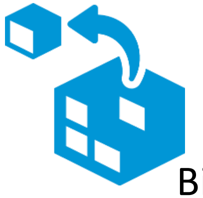
  - *20 applications loaded every second*
  - *2 resumes posted on Linkedin (200 million resumes) and …*
- one identity theft !

Big Data is the evolution of computing boundaries

INSTITUTE OF TECHNOLOGY

*one Zeta Bytes (ZB) = 10**21; 1000 EXA*

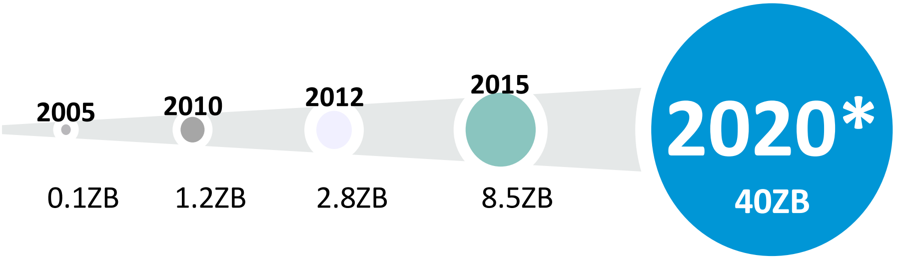
Cloud/IOT



Big Data

Mobility

Volume



Variety

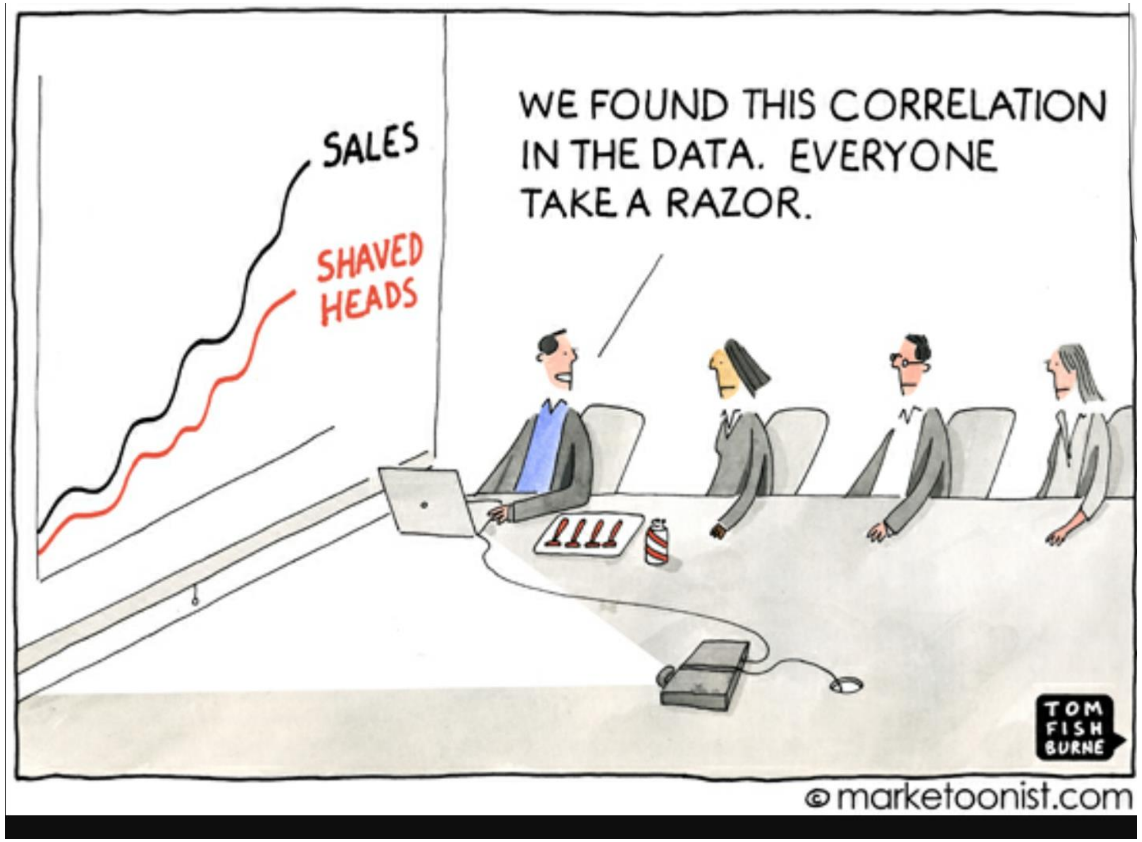
Velocity

IDC Estimates that by 2020, business transactions on the internet - business-to-business and business-to-consumer - will reach 450 billion per day.

*Source : IDC Digital Universe in 2020

« DATAFICATION » & CORRELATIONS

➢ CORRELATIONS (HOW) >> CAUSES (WHY)



The « 8V » of BIG DATA (2020)

the « 3 V » +

+ VALUE

+

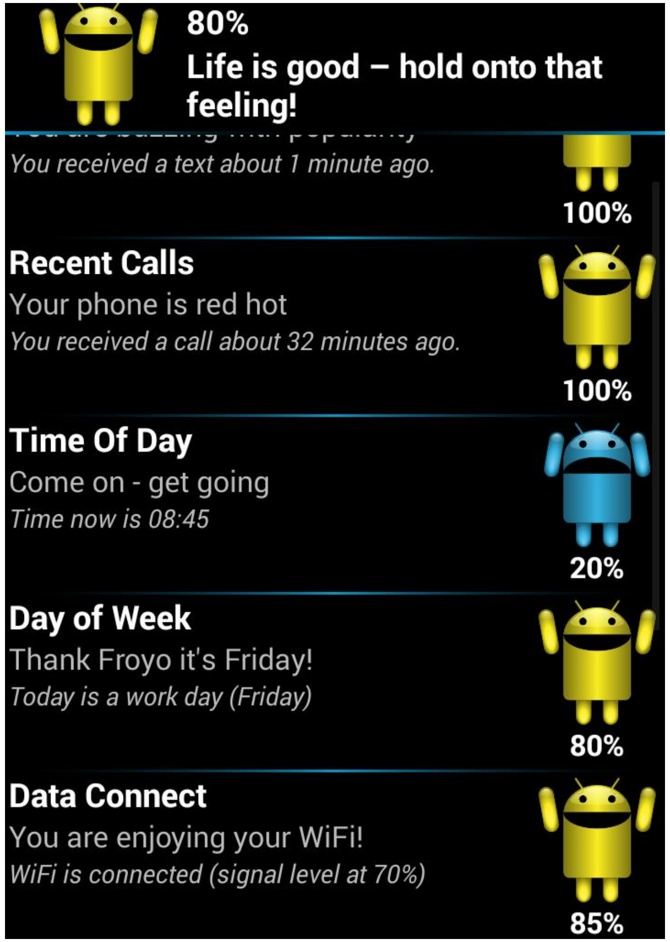
- Veracity
- Viscosity
- Visualisation
- Virality
- +++

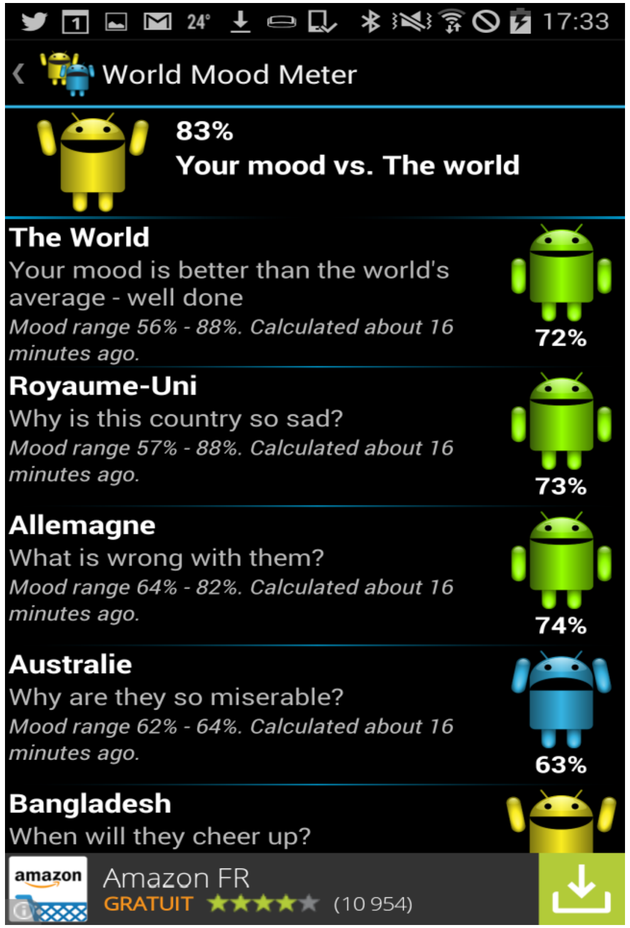
Extra « V » for VALUE : « Little » BIG DATA

INSTITUTE OF TECHNOLOGY

LIBERATION, 2014

VALUE : Ex Moodmeter





The DATA WAR on CONTENTS for our… « *digital clone/digital assistant* »

- BOOKS
- TV
- SPORTS (New comers : Amazon, Facebook, SFR,..)
- MOOCs (courses)

- DIGITAL RULE §
  - « ***If you do not pay for the product, YOU ARE THE PRODUCT*** ! »

CLOUD COMPUTING

« everything as a service (EaaS) : « SERVICE SCIENCE » (IBM, 2011) !

- « INFRASTRUCTURE as a SERVICE » (IaaS)
- « PLATFORM as a SERVICE » (Paas)
- « Software as a Service » (SaaS)

- DATA ? « DATA SCIENCE »
  - « **DATA as a Service** » Oracle (DaaS)
  - « **ANALYTICS as a SERVICE** » (AaaS) Google, IBM, Bigquery (Google 2012)

Parallelism and Big data

- **Terabytes (10**12) per second ?**

- Typical hard disk : 100 Megabytes/sec
  - 1 Terabytes (10**12)
  - 10 000++ hard disks in parallel

- 3 Solutions :
  - Data compression
  - SCALE UP : (SMP, MPP)
  - **SCALE OUT**

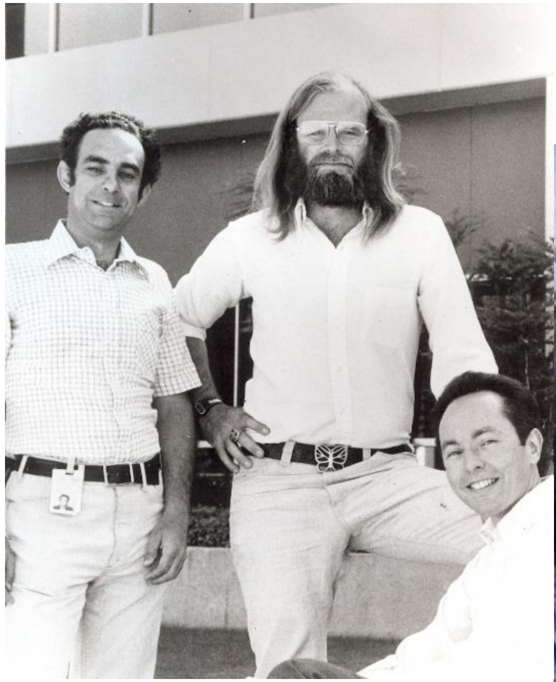
Three major modes of DATA processing (Jim Gray)

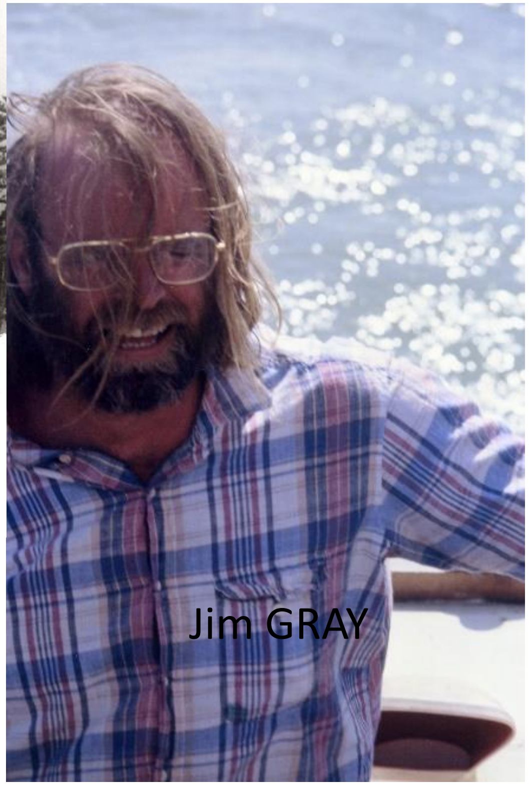
- DATA management (and storage/CLOUD)
  - DATA RETRIEVAL
    - READ focus : CODD's theorem
    - ➔ SQL esperanto
  - DATA UPDATE
    - DATA TRANSACTION (production support)
    - WRITE focus : GRAY's theorem
- DATA Analysis (decision support)
  - Data Learning (Machine learning & Deep Learning)
- DATA Communication

Gray's dream for… BIG DATA and his 4th paradigm of science (Turing Award 1998)

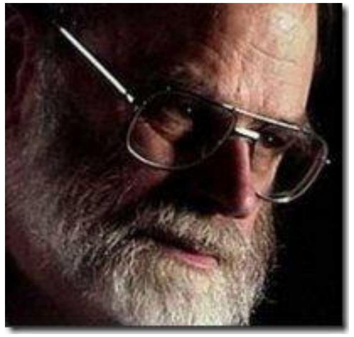
- *"to have a world in which all of the science literature is online, all of the science data is online, and they interoperate with each other."*

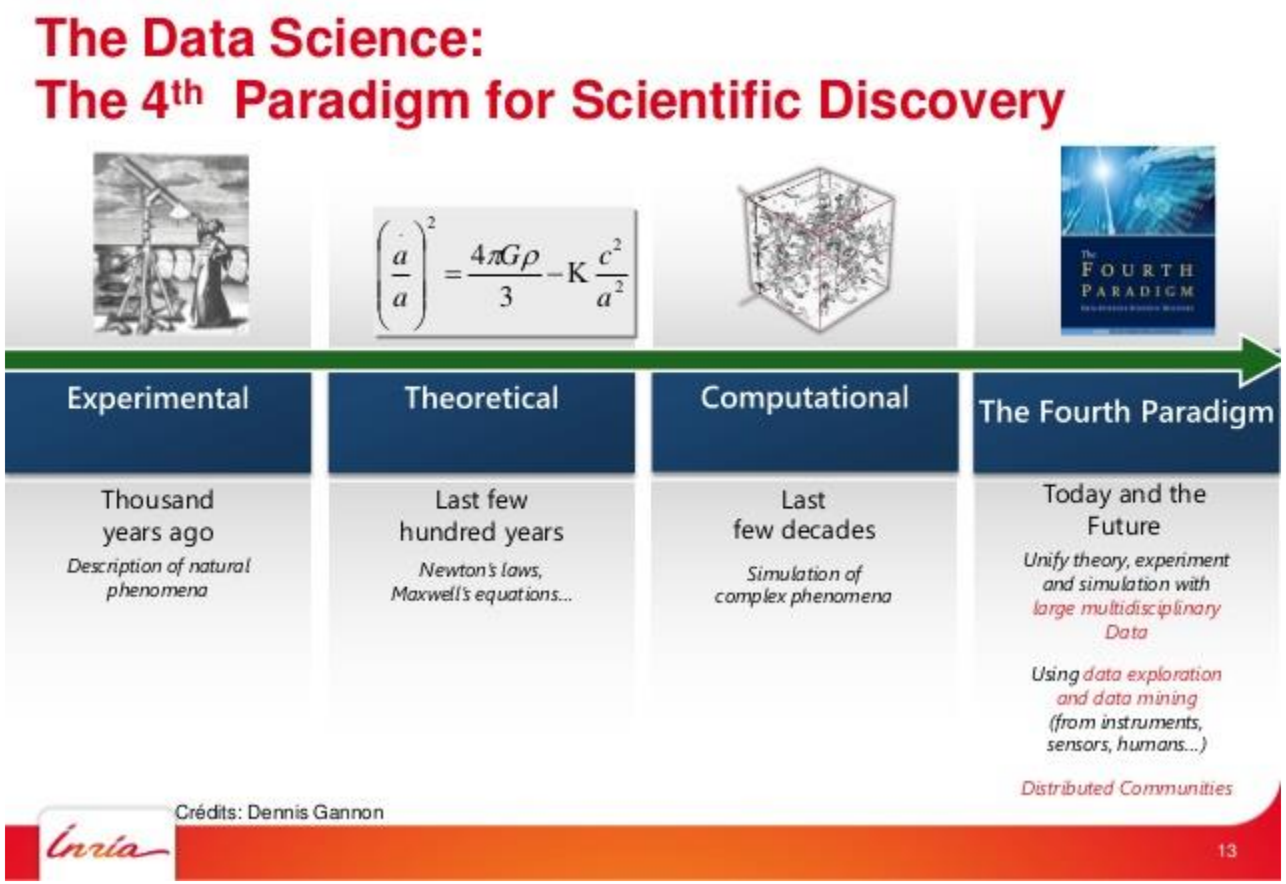
- Open source platforms for
  - DATA Management ➔ (Hadoop, Map Reduce, MongoDB,..)
  - DATA AnaLysis /Data Science ➔ (R language, Tensorflow, ..)
  - DATA Communication (Internet)





The 4th paradigm of science : the DATA paradigm !





BIG DATA : a couple in Sciences !

« BIG DATA is an ART crossing different sciences »

(CS, Maths)

BIG DATA in science ?
Computer Science & Mathematics

1. **DATA MANAGEMENT** (data-lake creation; data engineering) :

SQL3, OQL, BigquerySQL, NOSQL, CQL, HQL, HiveQL, N1QL, SPARQL, SPARKSQL, UnQL, coSQL, NEWSQL,..

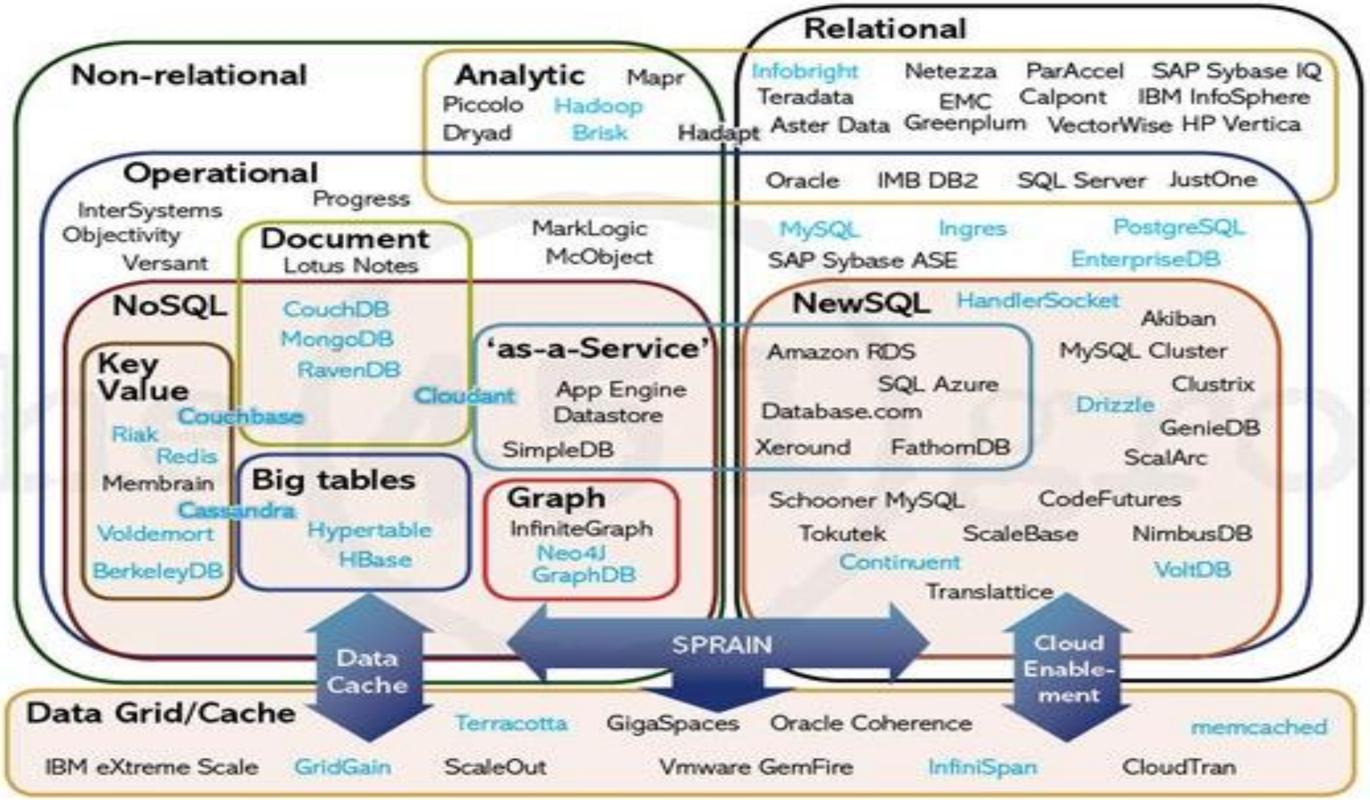
REF Open Source : HADOOP/MAP REDUCE, MongoDB, Cassandra…

1. **DATA ANALYSIS** (DATA Science; data mining)

Ref OPEN SOURCE :
R language (> 4000 packages), PYTHON, TENSORFLOW, CAFFE,…

Plethora of BIG-DATA management Systems (Aslett, 2015)

- https://blogs.the451group.com/information_management/2011/04/



BIG DATA management systems

- TOP DOWN approach for structured and semi-structured DATA
  - SQL2, SQL3, ODMG
  - Semantic Web (SPARQL, OWL)

- BOTTOM UP Approach for UNSTRUCTURED DATA
  - N.O. SQL (NOT ONLY SQL)
  - NEWSQL

DATA BASE market & standards?

PROCESSING

(Stonebracker 96 & Gartner)

	Simple	Complex (graphs,..)
SQL	(1) R-DBMS **SQL2** *Production Decision*	(2) OR-DBMS **SQL3** *Mobiquitous & Big Data systems*
No SQL	File System	(3) OO-DBMS **ODMG** *CAD*

DATA

	2010	2020
(1)	10 G dollars	20% of cont. Growing rate
(2)	2x (3)	**2x (1)!**
(3)	1/100 x (1)	1/100 x (1)

TOP DOWN approach for structured and semi-structured DATA BASES

- *TOP DOWN approach with* ***predefined SCHEMA and metadata***
- **STRUCTURED DATA standards**
  - SQL2 and **VALUE paradigm**
  - SQL3 and **POINTER-VALUE paradigm**
  - ODMG and **OBJECT-VALUE paradigm**
- **SEMI-STRUCTURED DATA standards**
  - SPARQL and **PREDICATE-VALUE (RDF)** paradigm

DB contibution to Computer Science : TIPS properties

T	**T**ransactions (with *ACID properties*)
I	No-procedural **I**nterface (SQL)
P	**P**ersistency (virtual paged memory)
S	**S**tructuration (Schema)

Object contributions to the DB World : RICE properties
an object = (OID, Value)

INSTITUTE OF TECHNOLOGY

- **R** Reusability (Inheritance or polymorphism)
- **I** Identification (OID : Object Identifier)
- **C** Complex Object construct
- **E** Encapsulation (Methods)

Semi-structured and unstructured DATA

W **Web DATA** (XML, RDF, Open Data, CSV, JS Documen

H **Hadoop Map Reduce** (Hive)

A **Analytics (ML & DL, ..)**

T **Time dimension (real Time); Tags; Timestamps**

DATA PARADIGMS

Codd's relational data model
(SET theory)
VALUE paradigm

***Predefined Schema* for STRUCTURED DATA**

OBJECT data model
(GRAPH THEORY)

TIPS

SQL2, SQL3/ODMG
NEW SQL
BigSQL

RICE

Big Data Management Systems

POINTER-VALUE paradigm
(SQL3)
OBJECT-VALUE paradigm
(ODMG)

N.O. SQL/
NEW SQL

***Meta data* for semi-structured data**

SPARQL
(OWL)

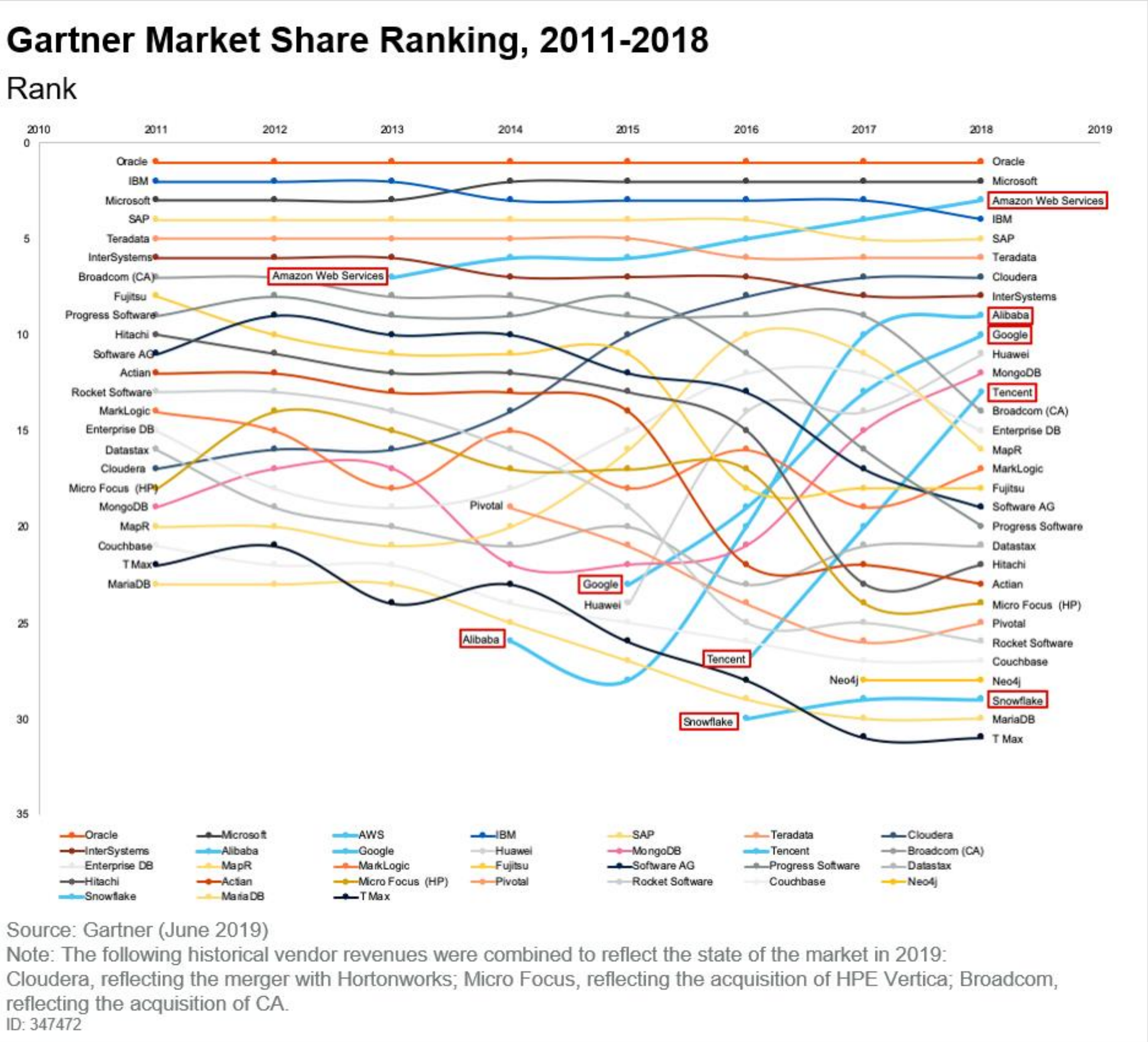
WHAT

No schema & NO Meta data for unstructured data

PREDICATE-VALUE (RDF) paradigm
(Semantic web)
(GRAPH THEORY)

KEY VALUE paradigm
(Map Reduce)
(MATRICES and linear algebra)

Gartner 's quadrants on DBMS (2019)

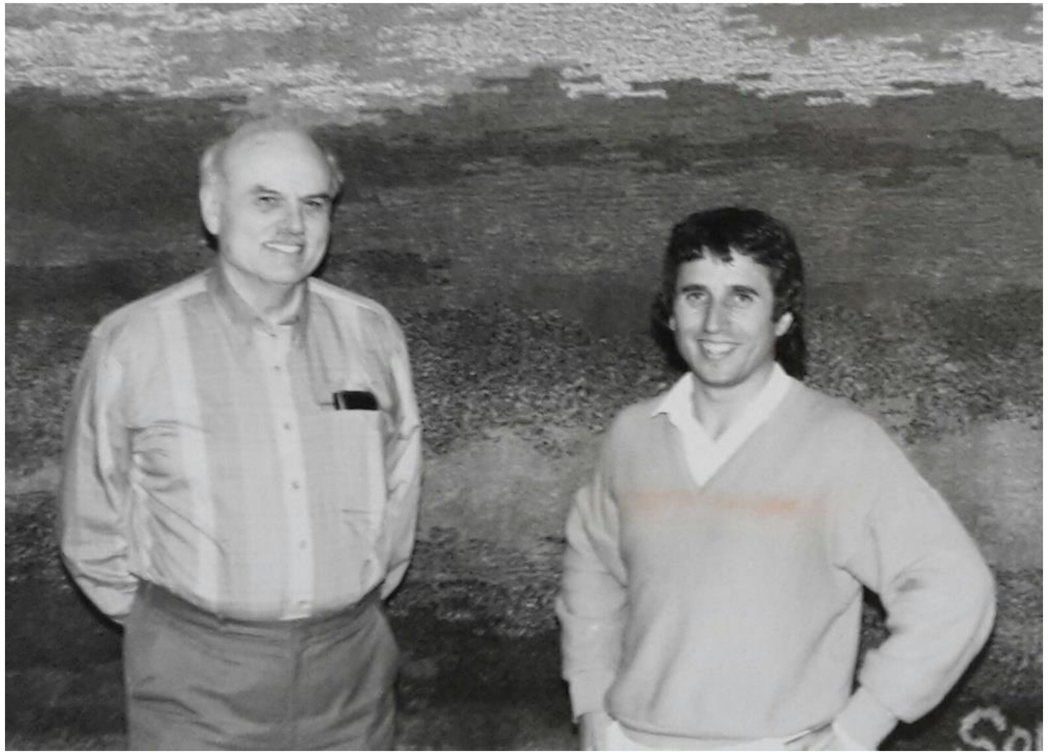


« The future of DBMS market is the CLOUD » (GARTNER)

- DBMS cloud services are already \$10.4 billion of the \$46.1 billion DBMS market in 2018.
- the overall DBMS Market grew at 18.4% from 2017 to 2018 – its best growth in over a decade. Cloud DBMS accounted for 68% of that growth.
- Only two vendors (Amazon Web Services and Microsoft) account for 75% of the growth from 2017 to 2018. AWS is 100% cloud and Microsoft DBMS was almost 100% cloud (See previous Figure).

- **DBMS innovation is cloud-first or cloud-only for development**

BIG DATA MANAGEMENT A.C. (After Codd)



With Ted Codd in Sophia Antipolis (1986)

- **Ted CODD 1968 (Relational ALGEBRA & SQL2) :**
  « *Everything is VALUE* »

- Chris DATE & Mike Stonebraker 1995 (SQL3) :
  « ***Everything is POINTER-VALUE*** »
- TIM BERNERS LEE 1998 (SparQL, RDF) :
  « ***Everything is PREDICATE-VALUE*** » ***(WEB DATA)***

- Chang 2006 (N.O.SQL) :
  «***Everything is KEY-VALUE*** »
- STONEBRAKER 2013 (NEW SQL) :
  « ***Everything is SQL*** »

- ElMore 2015 « ***Everything is POLYSTORE*** »
- Deep Learning : ***« Everything is an IMAGE »***

- and Evariste Gallois 1832 !! :
  « ***Everything..is a GROUP (CATEGORY)*** »

- **Jim Gray :** « ***Everything is DATA (4th paradigm of science)*** » ***…and*** « ***Everything is a TRANSACTION*** »

Bottom up approach for unstructured data (no schema, no metadata)

- « N.O. SQL » (Not Only SQL)
  < meaning NO Relational>
  - « KEY /VALUE Paradigm »
  - GRAPH paradigm

- « NEW SQL »
  - « SQL paradigm »

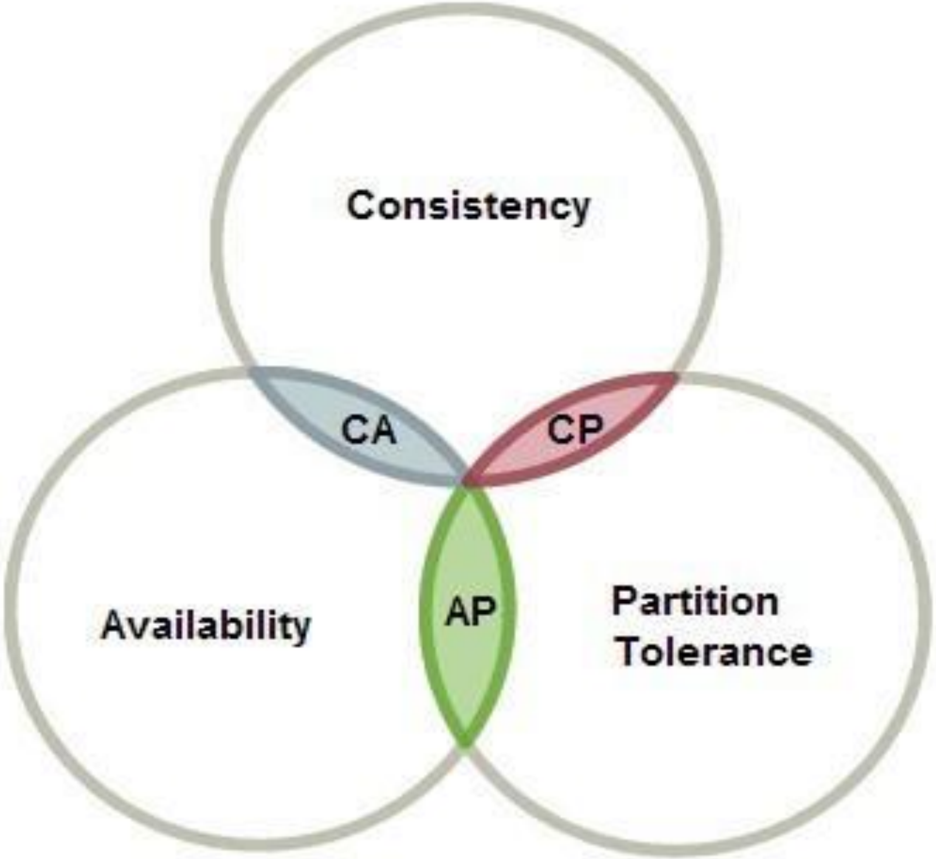
« BASE » properties

- BASE :
  - Basically
  - Available
  - Scalable (OUT)
  - Eventually consistent (final consistency)
    - *Replica consistency ; Cross Node Consistency*

- CAP Theorem

  (Eric Brewer, Prof Berkeley, 2000 & 2012 ; Revised by Altend MIT, 2002)
  - Consistency, SQL
  - Availability,
  - Partitioning NO SQL

CAP Theorem : « Pick 2 ! » (Brewer 2000 ; 2012)



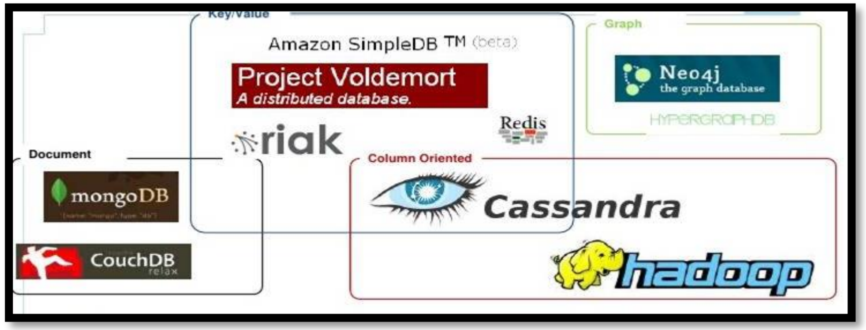
NOT ONLY N.O. SQL (Not Only SQL)

(1998)

4 « no » :

1. **no SCHEMA** (schema-less ; *Variability*) **& NO METADATA**
2. **no RELATIONAL/ NO JOIN** (extract data without joins)
3. **no DATA FORMAT**(graph, document, row, column)
4. **no (ACID)Transactions** *(CAP theorem ; BASE)*

+ (VALUE)… + VOLUME + VELOCITY + VARIETY



2 Complementary approaches for big data management

	SQL	N.O.SQL
VOLUME & VARIETY	STRUCTURED (SCHEMA) TERA/PETA bytes	Unstructured (no schema) EXA/ZETA++ bytes
VELOCITY	NO	YES
TRANSACTIONS	YES (ACID and Gray's theorem)	NO (BASE & CAP theorem)
SCALABILITY	UP (Scale up)	OUT (scale OUT)
USER INTERFACE	AD HOC Queries, JOIN & Transaction oriented	Predefined queries, NO JOIN & Decision oriented
STANDARDS	SQL3/ODMG	Not yet (BIG SQL)
Typical approach	TOP DOWN (predefined Schema)	Bottom UP (no schema)
Administrator	Yes	No
Vendor support	Yes	No (Open Source)

« N.O. SQL » DBMS

4 data paradigms: 3 KEY-VALUE oriented and one GRAPH oriented

- KEY-VALUE with BLOBS (Binary Large Objects)

  ex : **Hadoop**, **Cassandra**, Ryak, Redis,DynamoDB, BerkeleyDB, etc.

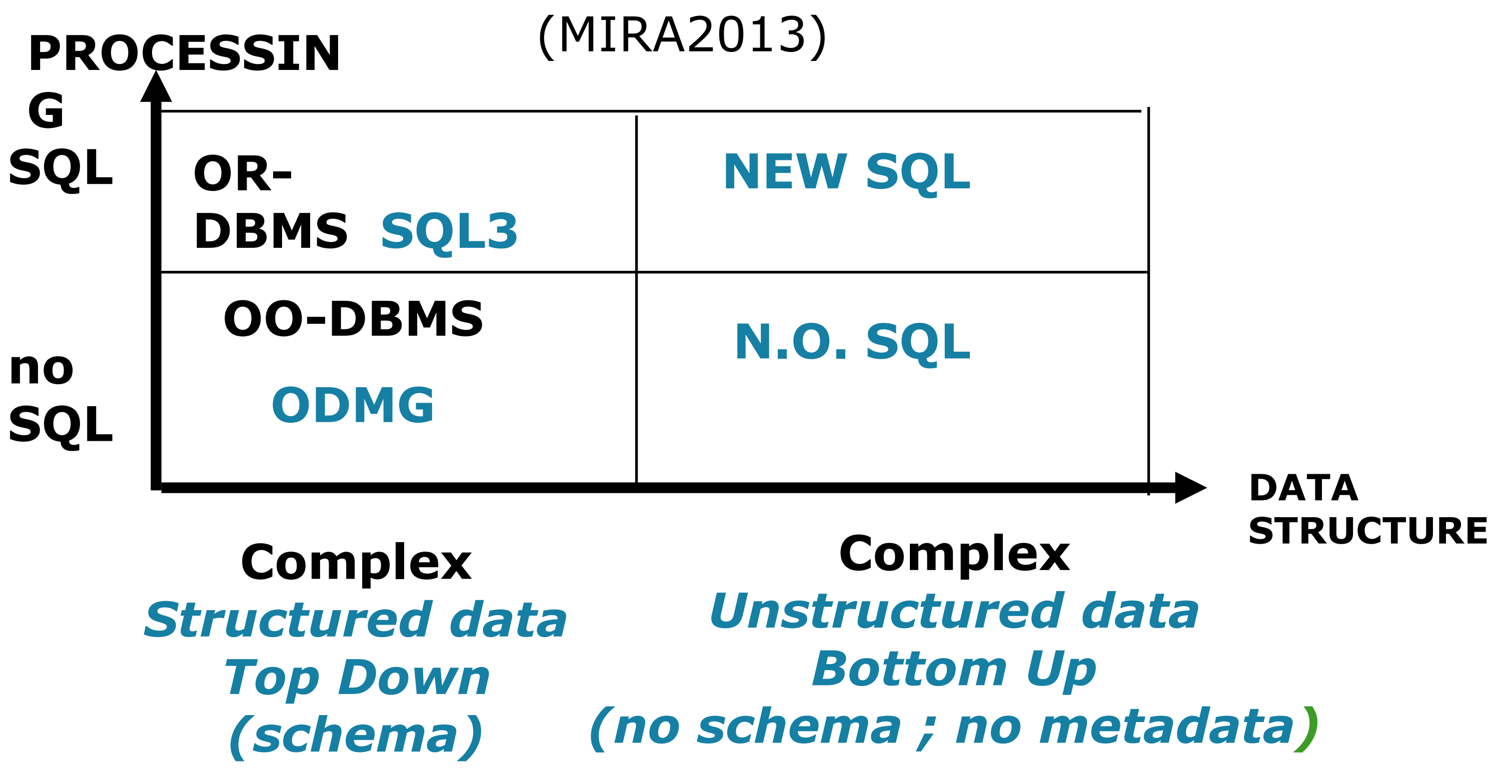
  → **HASHING arrays (no query engine)**
- KEY-VALUE with JSON/XML documents

  ex : **MongoDB**, **CouchDB**, etc.
  - *JSON simpler than XML with Java Script interface*
  - <KEY, VALUE> model with VALUE in JSON (BSON, XML) for documents ;
- KEY–VALUE with COLUMNS

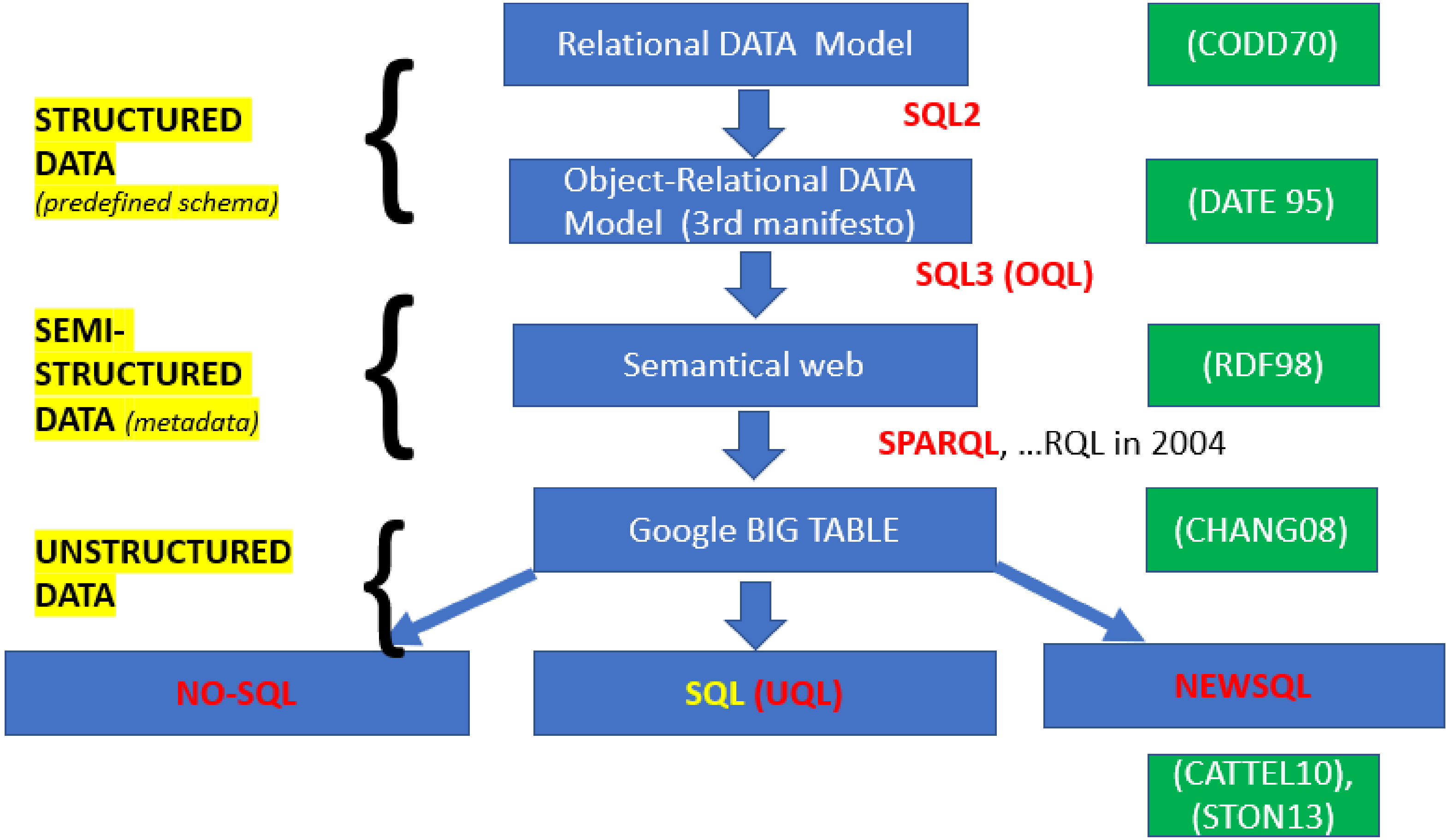
  ex : **HBASE**, **Cassandra**, BigTable/Google,…
  - <**KEY**, (SETofcolumns, VALUE, TIMESTAMP)>
- GRAPH oriented

  ex: **Neo4j**, OrientDB… : towards GQL (Graph Query Language)

« COMPLEX » data : SQL3, N.O. SQL et NEWSQL?



SQL evolution (a.c.)



A relational-schema example with three predicates

PILOT (<u>PIL#</u>, PILNAME, ADDR)

PIL#: Pilot ID then NAME and ADDRESS (City)

PLANE (<u>P#</u>, PNAME, CAP, LOC)

CAP : Capacity, LOC : localization city

FLIGHT (<u>FL#</u>, PIL#, P#, DC,AC, DT, AT)

DC : Departure City, AC : arrival city,

DT : Departure Time, AT : Arrival Time

Note :

Primary keys are underlined

Example : SQL2 (Relational)

Who are the pilots (PIL#, PILNAME) from Nice driving a plane from Nice ?

In SQL2:

SELECT PIL#, PILNAME

FROM PILOT, FLIGHT

WHERE PILOT.PIL#= FLIGHT.PIL# and PILOT.ADDR= 'Nice' and FLIGHT.DC= 'Nice';

In Codd's relational algebra :

V1 = Join PILOT (PIL#= PIL#) FLIGHT

V2 = Select V1 (ADDR= 'Nice' and DC='Nice')

RESULT = Project V2 (PIL#, PILNAME)

Example : SQL3 (object relational)

Who are the pilots from Nice driving a plane from Nice ?

In SQL3 :

SELECT REFPIL ➔ PIL#,PILNAME

FROM FLIGHT

WHERE DC= ‘Nice’ and REFPIL ➔ ADDR =‘Nice’;

Note : with

- *REFPIL : REF type attribute containing ROWID (OID) from PILOT and*
- « ➔ » : *Dereferencing operator*

Example : OQL (ODMG)

Who are the pilots from Nice driving a plane from Nice ?

In OQL

```
SELECT p.PIL#, p.PILNAME
FROM
        p in PILOT
        v in p.performFLIGHT
WHERE
        p.ADDR= 'Nice ' and v.DC='Nice';
```

Note : *with « performFLIGHT », bidirectional persistent REF pointer from PILOT class towards FLIGHT class*

SPARQL (Example)

Who are the pilots from Nice driving a plane from Nice ?

Prefix rdf :<http:// www….>

SELECT ? PILOT

WHERE { GRAPH ?g

{ ?PILOT rdf :ADDR rdf: Nice

?FLIGHT rdf:DC rdf: Nice }}

Some SQL examples on KEY-VALUE NO SQL DBMS with N1QL (Couchbase), CQL (Cassandra), …

Typical Example :

N1QL :

SELECT PIL#, PILName
From Pilot
Where ANY F in Flight SATISFIES F.DC= ‘Nice’ and ADDR = ‘Nice’;

CQL3

SELECT PIL#, PILNAME
From PILOT
JOIN EACH FLIGHT ON Pilot.pl#=FLIGHT.PL#
and DC =‘Nice’ and ADDR = ‘Nice’;

Example : GQL (Graph Query Language)

➢ **Query : Name of the pilots who perform a flight from Nice with a Boeing 747 ?**

```
SELECT
pl.name AS pilot.name
FROM GRAPH pilotflightsplanes
MATCH
 // graph pattern
(pl:Pilot)->{:PERFORMS}->(f:flight)<-{:IS-USED}<- (p:plane)
WHERE
    p.name = ’B747’ and f.DC = “Nice”;
```

- The pattern means that all data in the graph that matches the sequence of nodes and edges (each of which has a particular « label » or element type) will be identified.
- This operation lifts a «sub-graph » or a « projected graph » of flights for a particular pilot into the application.
- Properties on all instances of :PILOTnodes or :FLIGHTedges that match can now be read by the application.

Example : JSON integration (new data type in SQL)

- ***Postgres (2013), Oracle (2013), SQL Server (2016)***
  - New data TYPE containing Json documents : **Jsonb**
    - Examples :

      Create Pilot (PIL# : integer primary key, , Pilname : varchar (12), **flightreport : Jsonb)**
      Select Jsonb_each (flightreport) from pilot;
  - JSON view creation for SQL data (and conversely)
- Example : CREATE JSON_VIEW AS

  SELECT JSON *{"pilot" : {"pil#" : pl.pil_id, "Pilname" : pl.name, {"flight" : {"f#" :f.flight_id, "DC" : f.city, "AC" : f.city}} }*
  FROM *pilot pl, flight f*
  WHERE *pl.pil_id = f.pil_id* ;
- *With **functors** (categories) <see later>*

Towards « BIG SQL » standard

SELECT

FROM

{T/Tables}, {V/Views} < SQL2; Create Table..Create View,..)
{ SQL query} < SQL3> JOIN
{Pointer Chasing} <OQL>
{**EXTERNAL TABLES**} < N.O.SQL DB>
< Oracle, IBM, Microsoft, Informix, Sybase/SAP, MySQL,..>
{**GRAPH**} < GQL>

WHERE

<POINTER DEREFERENCING operator on REF types> <SQL3>
<Multiset operators> like **NEST/UNNEST** <Multivalued attributes>
<**GRAPHS** operators>, <**MATRICES** operators>, UDF/UDT, **MAP/REDUCE**

PARTITION BY <Splitting/Sharding>like **LIMIT/OFFSET, PIVOT/UNPIVOT**

MATCH <GQL>

GROUP BY /HAVING

GROUPING SETS with **CUBE, ML & DL operators**

Looking for a unifying theory for BIG DATA management

« An effective mathematical model that encompasses the concepts of SQL, NEW SQL, NO SQL, (ML and DL) which would enable their interoperability » J. KEPNER (MIT, 2016)

- **4 types of DATA VARIETY with 3 theoretical frameworks and one core theory**
  - **STRUCTURED** (SQL) and **SET theory** (of VALUES/DATA) < core theory>
  - **SEMI STRUCTURED** (SparQL, OWL) : **GRAPH THEORY** (inferences)
  - **UNSTRUCTURED (NOSQL)** : **GRAPH THEORY**
  - **NEW SQL** and **MATRICES** (linear algebra)
    - NOTE : **DATA SCIENCE (ML and DL) and MATRICES management**
- **Specific extensions of set theory and SQL to tackle :**
  - **Hierarchical data: JSON, XML**
  - **Graph data : RDF, OBJECTS, Social networks (and GQL)**
  - **Tabular data : Tables, CSV, OLAP, tensors, pixel array (Deep Learning)**

Tribute to an *assassinated* revolutionary mathematician and its contribution tothe future of DATA modeling for BIG DATA

INSTITUTE OF TECHNOLOGY

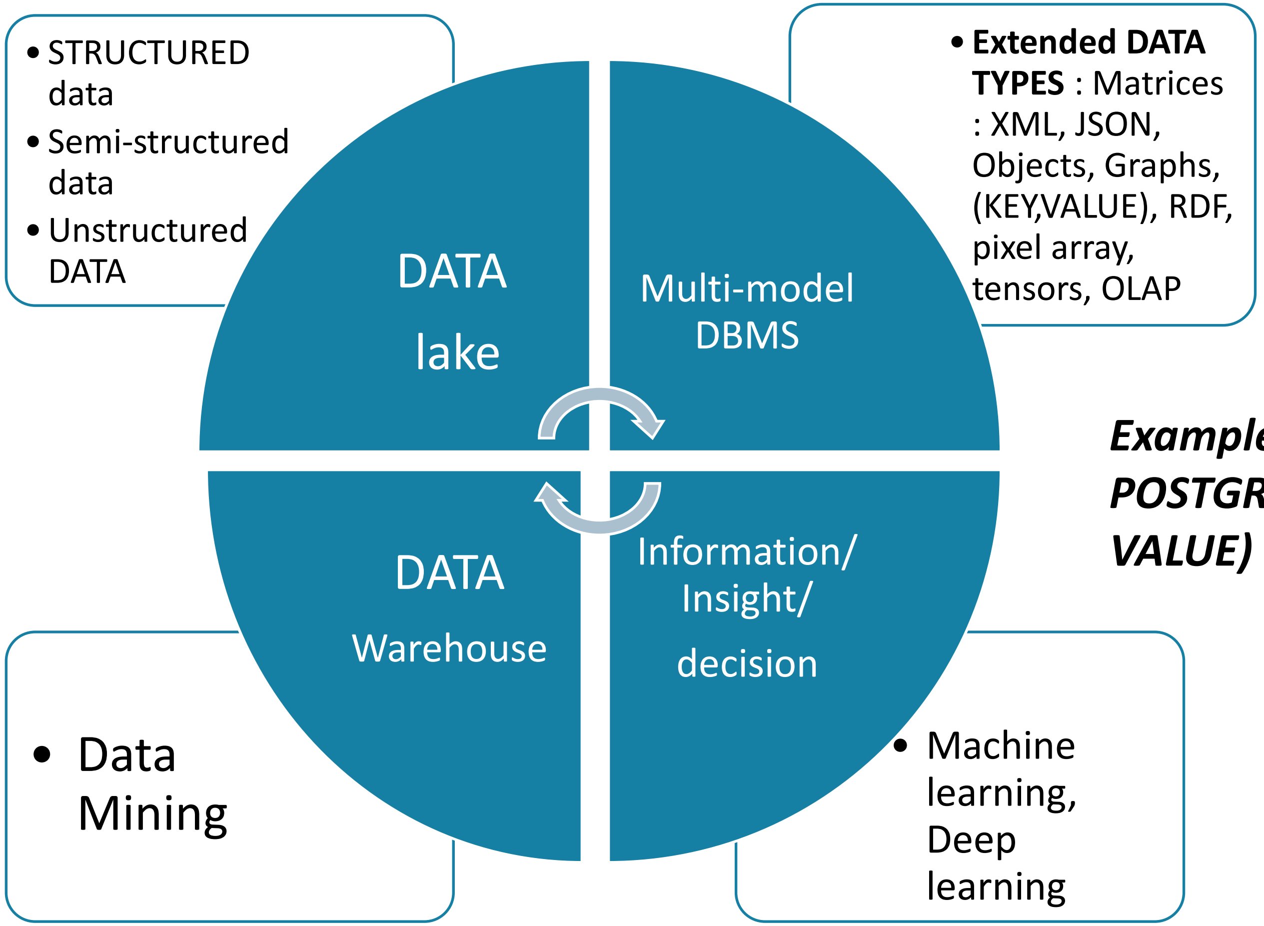
Evariste GALOIS : (1811- « *killed in a duel* » in 1832 when he was 21 years old) :

Letter to Auguste Chevalier, 29th of May 1832 (eve of his death)

« GROUP theory »

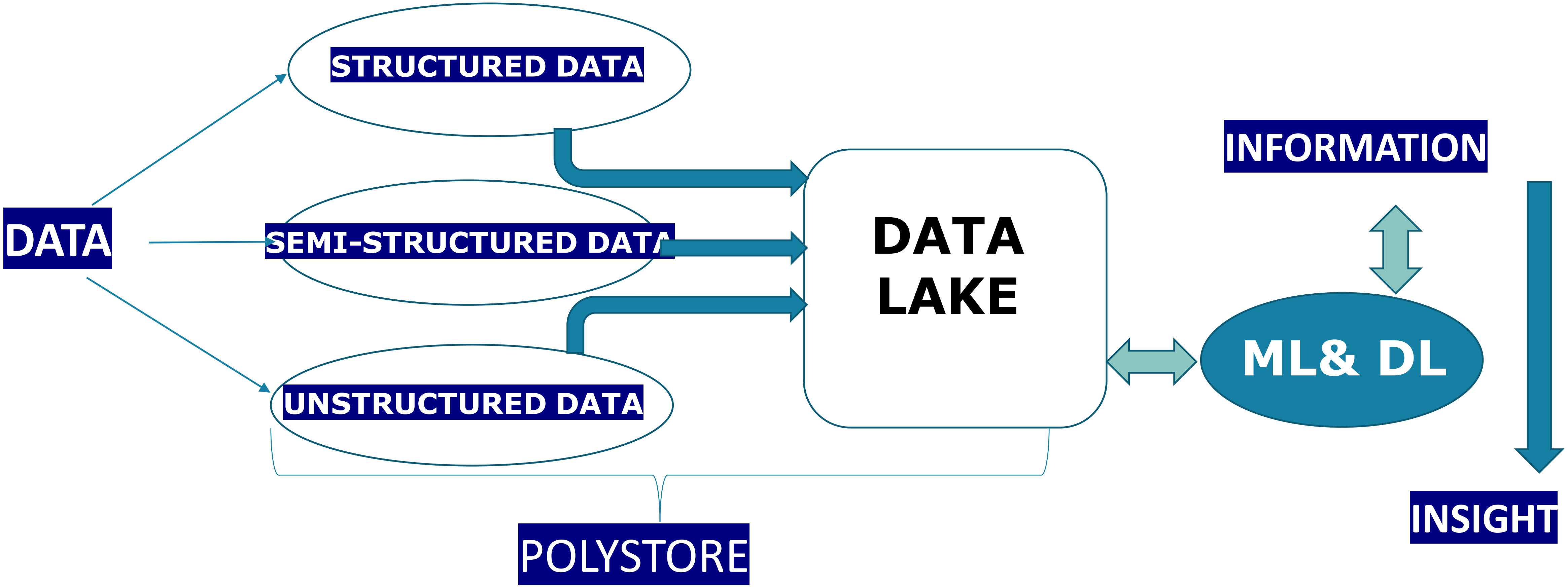
→ « Category theory » and « Associative arrays » !

DATA refinery



Example : Multi-model POSTGRES : SQL+ +NOSQL (Key, VALUE) + JSON + XML in 2019

DATA LAKE and polystore



Dual approach for data lakes : polystores* vs Multi-model data store

« *All leading operating DBMS will offer multiple data models, relational and NO SQL in a single DBMS platform* » GARTNER 2016

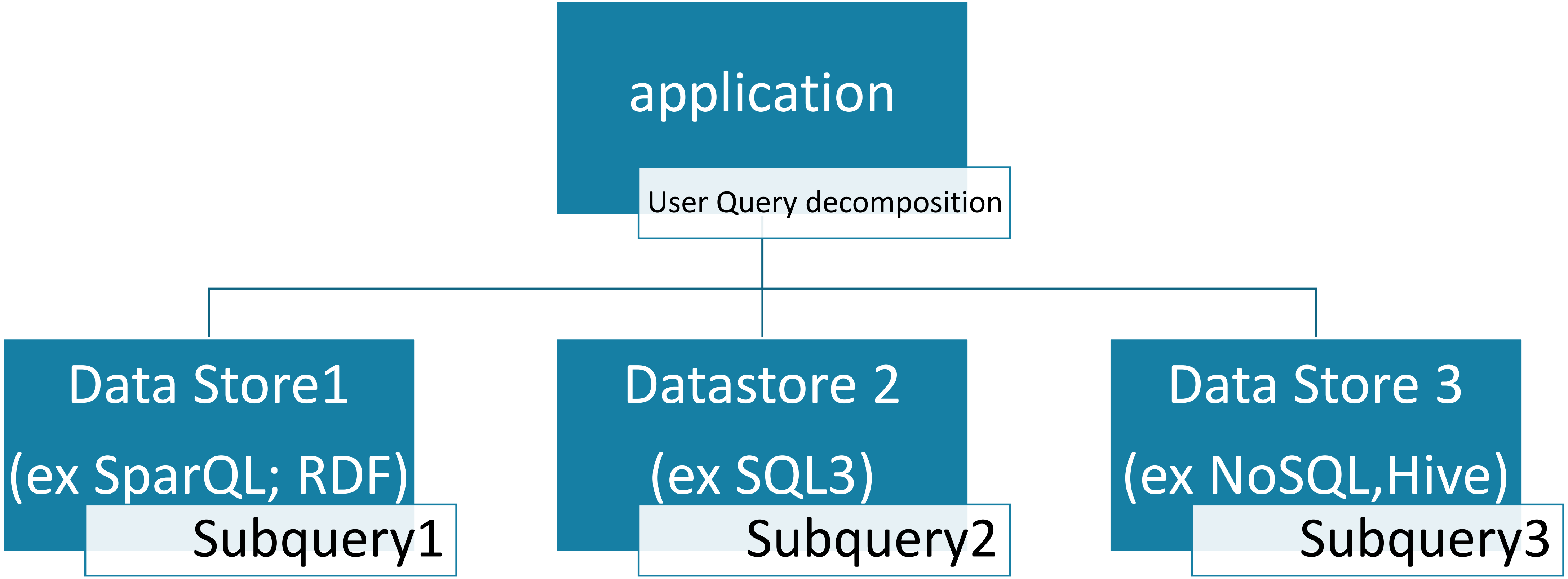
« *One size cannot fit all* » (M.Stonebraker)..but can « *fit a bunch* » (Asterix DB)

- Goal : one application accessing multi-model DATA sources ! With 2 solutions : POLYSTORES vs MULTI-MODEL DBMS
- **POLYSTORES* : integration of multiple autonomous data stores**
  - EX : Polybase (Microsoft); SparkSQL, BigDAWG (MIT)
  - 3 types ** :
    - Loosely-coupled (common interface, mediator/wrapper around autonomous systems)
    - Tightly coupled
    - hybrid
- **MULTI-MODEL DBMS : one single integrated DBMS around a unified data model**
  - EX Oracle, ORIENTDB, IBM, Cache, ARANGODB ..;
  - New types : sets, graph, key-value, object, Json, XML, RDF, text, multimedia, spatial, time-series, matrices, etc.
- ***➔ Need for a unified model theory for multi-model DB***

Mike Stonebraker « The case for POLYSTORES » 2015, http://wp.sigmod.org/?p=1629

** Bondiombouy, Carlyna, and Patrick Valduriez. "Query processing in multistore systems: an overview." International Journal of Cloud Computing 5.4 (2016): 309-346

Polystore



Examples with Oracle : Oracle MySQL, Oracle DB (Relational, Json, XML) and Oracle NO SQL (relational and (KEY, VALUE))

Bridge between
SQL, NoSQL and NewSQL ?

- Bridge between
  - SET ➔ structured /SQL
  - GRAPH ➔ semi-structured & unstructured (NoSQL)
  - MATRIX ➔ NewSQL & Data Analysis/R & Python

- Uniform Mathematical model for SETS, GRAPHS and MATRICES + JSON, XML, OLAP, OBJECTS ?

Two formal unifying BIG DATA management

- **CATEGORY THEORY** ***: Microsoft research <2011> and Oracle (+ Univ Helsinki) <2018>***

  - « ***A co-Relational Model of Data for Large Shared Data Banks*** », Erik Meijer and Gavin Bierman *Microsoft Research, CACM 2011*
    - M. Fokkinga, *« SQL versus coSQL — a compendium to Erik Meijer's paper » 2012*

  - **UDBMS: Road to Unification for Multi-model Data Management** Jiaheng Lu 1, Zhen Hua Liu 2, Pengfei Xu 1, Chao Zhang (1-University of Helsinki, Finland and 2-Oracle, Redwood Shore, CA, USA), *Dec 2016*
  - ***"Multi model databases and highly integrated polystores"*** J.Lu et al , *Tutorial CIKM 2018*

- ***ASSOCIATIVE ARRAYS : MIT < 2015>***
  - Jeremy Kepner and al *: « Associative array model of SQL, NoSQL and NewSQL Data bases »* <MIT CS and AI laboratory, 2016>

  - Kepner, J. Chaidez, V. Gadepally, and H. Jansen, *« Associative arrays : Unified mathematics for spreadsheets, databases, matrices, and graphs »* arXiv preprint arXiv:1501.05709, 2015.

DATA SCIENCE ? BIG CAPTA !

INSTITUTE OF TECHNOLOGY

« DATA SCIENTIST » :

« Sexiest Job of this millenium » Harvard Business review

- Skills shortage: The Data Scientist Opportunity !
  - Top 10 Jobs that didn't exist 10 years ago (Linkedin 2014)
  - Data scientist is #5 !

« Without data you're just a person with an opinion. » W. Edwards Deming

« Computer Science » evolution ?

- « **SERVICE SCIENCE** » (IBM, 2011) & CLOUD COMPUTING

- « **DATA SCIENCE** »
  - Machine learning
    - **Deep Learning**

- NOTE :
  KAGGLE web site for data scientists, « *place to discover and seamlessly analyze publicly available data* »

Artificial intelligence ? Intelligence ?

AI and Deep Neural Net-DNN- is back with GPU computing power and BIG DATA

DNN Application domains : computer vision, image analysis, assisted car driving, natural language translation (Ex : DEEPL*), LIPNET (Oxford then Google), voice cloning (LYREBIRD) , robots (Ex MANTIS), Deep Fakes, BIG DREAM (Google) for enriched pictures..

**DeepL* translated an MIT book (800 pages) on… Deep Learning in 12 hours (from English to French) in October 2018

« Will Machines Eliminate Us? » Will Knight, MIT Technology Review, 29 January 2016

« AI paradox » and the data economy

ESTIA INSTITUTE OF TECHNOLOGY

- "***AI will contribute 13 TERA DOLLARS (****1 TERA : $1 trillion: 10**12)* ***to the global economy over the next 10 years*** (Harvard Business Review, Dec 2019, P.40).
  - AI is not an off-the-shelf technology
  - Moving from an experience-based TOP DOWN decision process to a DATA-based BOTTOM UP decision process
- "***« AI is colonizing every economic sector of our society : health, transport, industry and education** …**The prodigious expansion of AI is linked to deep learning, which allows to train a machine to perform a task without explicitly programming it … Deep Learning is the most promising form of AI because it is powered by the BIG DATA and GPU***"

  Yann Le Can, Turing award 2019

- *"We don't have better algorithms, we just have more DATA ."*
  Peter Norvig, current Google Head of AI (2019)

- **AI paradox** " Yan Le Can
  - **no "common sense"** or consciousness

    Highly specialized intelligence and not generalist (recognize tumor ok but unable to make coffee)

    Artificial intelligence is VERTICAL (highly specialized), human intelligence is HORIZONTAL (culture).

e-BIMAR

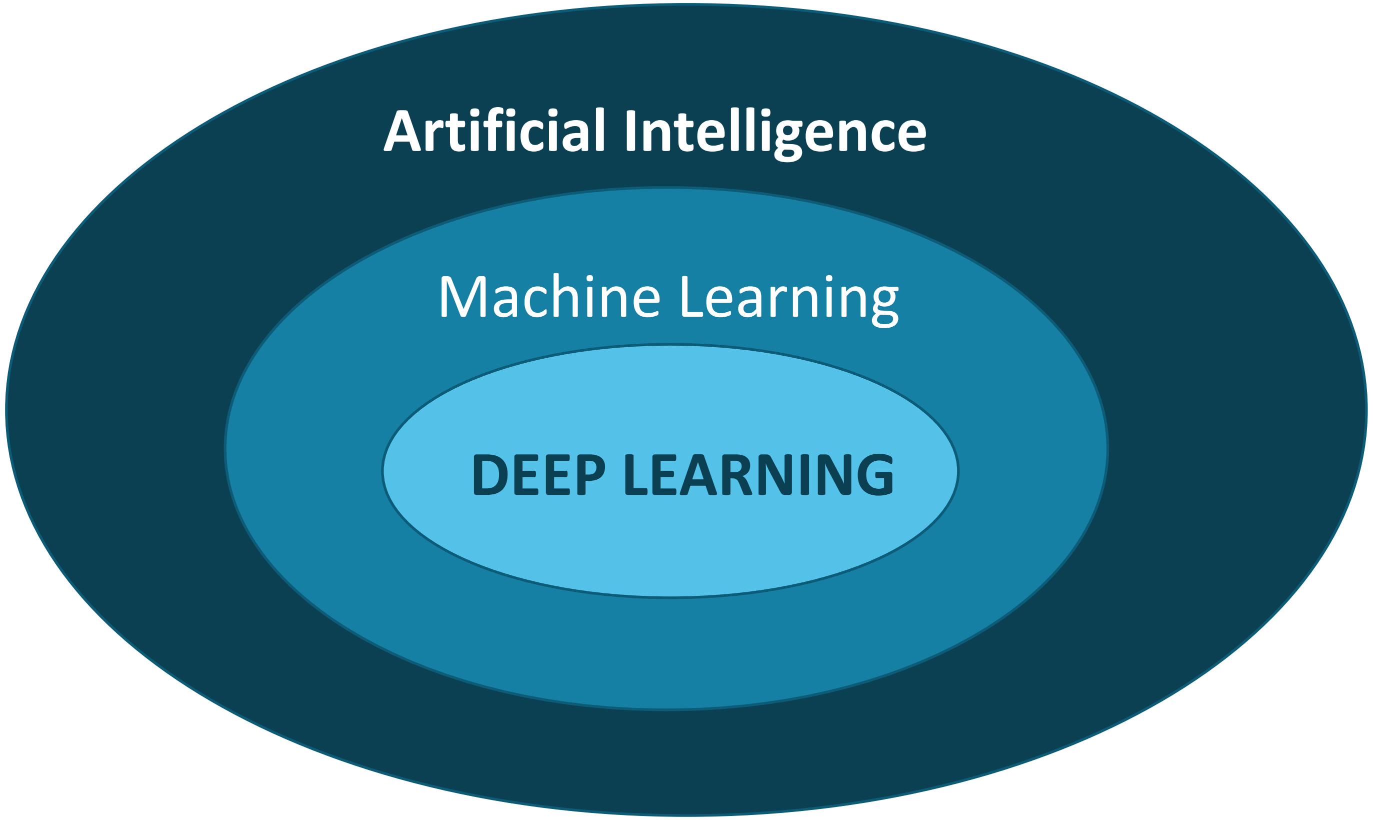
Some examples of recent gaming AI milestones

- May 1997 : DEEP BLUE (Ibm) beats Garry Kasparov, Chess World Champion

- 2015 : AlphaGO (GOOGLE) beats the GO World champion

- 2017 : LIBRATUS (Carnegie Mellon) won a poker marathon (1 766 250 dollars) against four poker world champions in Las Vegas (« Heads Up (1 vs. 1) No-Limit Texas Hold’em’ »).

DATA SCIENCE : AI, Machine Learning, Deep Learning

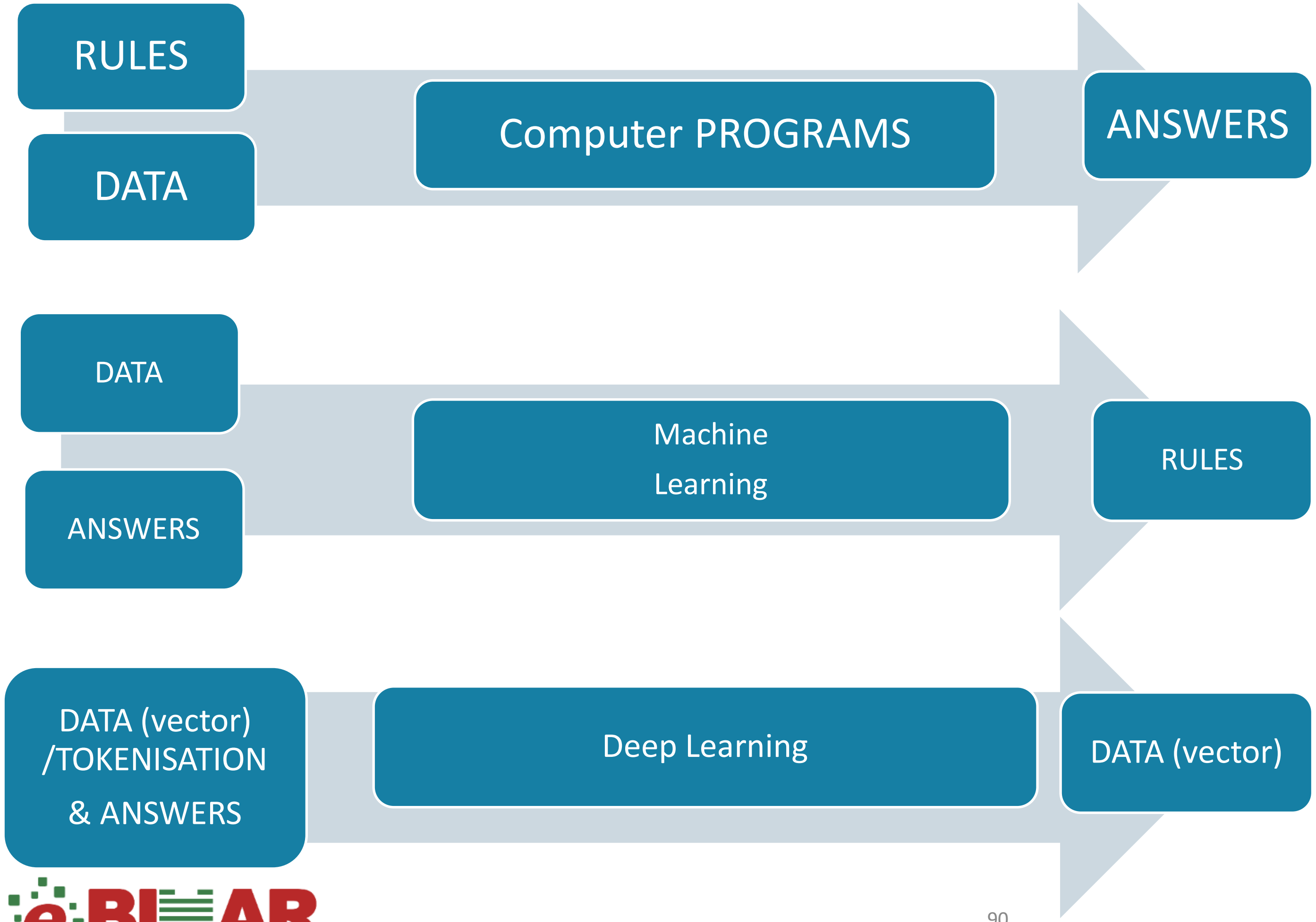


Machine Learning :
Machine capable to learn without being programmed for

Weak AI : IMITATION and TOP-DOWN

Strong AI : CREATION and BOTTOM UP

Symbolic AI , ML & DL



ML is trained with labelled DATA without explicit programming

ML is strongly associated with applied statistics but on BIG DATA

DL is a simplistic modeling of brain neurons

« Machine learning (ML) » (DATA driven):

applied statistics meeting BIG DATA!

INSTITUTE OF TECHNOLOGY

« No free lunch theorem »
(algorithm agnostic)

« Every model is false but some could be useful »
George BOX

- ML Description with 2 types of variables and 2 functions :
  - **Y= F(xi) , F real prediction function**
    - Xi : predictive variables (observation)
    - y : target variable (for prediction)
  - f « **predictive function** » (close to F)

- **Machine LEARNING ?**
  - **SUPERVISED** (data labelling with Y being either a VALUE or a CATEGORY) :
    - **REGRESSION**
    - **CLASSIFICATION**
  - **UNSUPERVISED** (clustering); No Y ! :
    - **CLUSTERING**
    - **ASSOCIATION**
  - **Reinforcement :** *Learning paradigm by trial-and-error, from rewards or punishments*
  - *Adversarial,...*

eBIMAR

Artificial Intelligence ?
= DIGITAL ALGORITHMS ?

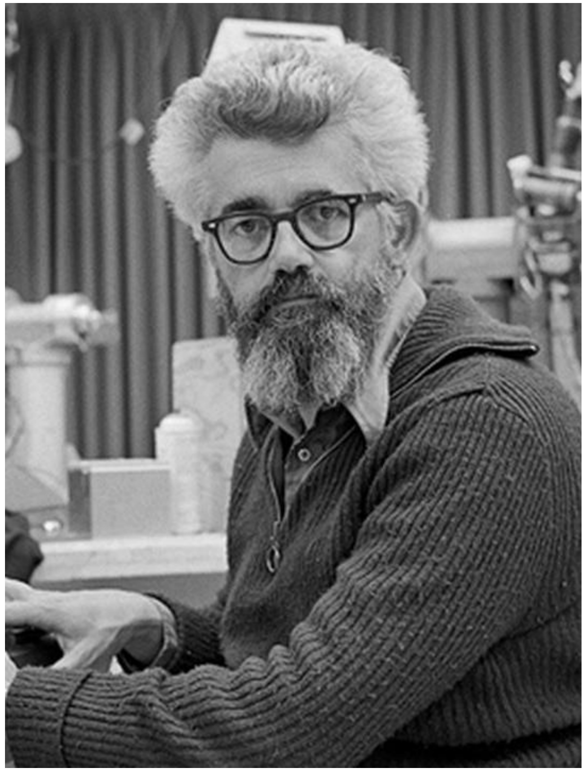
➢ **Born in 1957 after ..**
https://youtu.be/ZtwgqpUibfU

Dartmouth conference in August 1956 organized by Marvin Minsky and John McCarthy (MIT Professors …followed by « ***Moon project*** » John Kennedy and ….two long darkness winters …until 2012….

until BIG DATA and GPU !

2019 : TURING AWARD for *Y.Bengio, G. Hinton and Y. Le Cun < DEEP LEARNING>*

Artificial Intelligence ? not a (single) TECHNOLOGY !



➢ **a platform for DIGITAL ALGORITHMS** based upon different technologies : predicate logic, linear algebra, graph theory, …

with 2 generic formal approaches :
UNIVERSAL THEORY (brain model) vs EMPIRISM (Neural Nets,..)

SYMBOLIC AI vs CONNECTIONIST AI

In 2010 NEURAL NETS rebirth with BIG DATA and SILICON for AI (GPU)

INSTITUTE OF TECHNOLOGY

- ***SILICON for AI : CPU, ASIC and GPU***

- **CPU** *: highly programmable but no performance*

- **ASIC (***Application Specific Integrated Circuit)* and **ASSP** *(application specific standard product)* in Deep Learning
  - Special purpose integrated circuit (since the 80's)
    - Example : Matrix product
  - Training of neural networks for AI
    - Example : CLOUT (Google, 2017) Tensor Flow Processing Unit (TPU)

- **GPU** (*Graphical processing units*) **: Asic used for processing graphics!**
  - introduced to accelerate 3D graphics, gaming and video
  - GPU was then designed to perform MATRIX operation as fast as possible
  - High-level programming languages are compiled to the GPU (Java, Python, Matlab, Haskell, ..)

« SMART CHIPS » for DNN from CPU to GPU : « Moore's law is obsolete » ! JH Huang (Nvidia)

1965 : Intel founder « Moore's law »

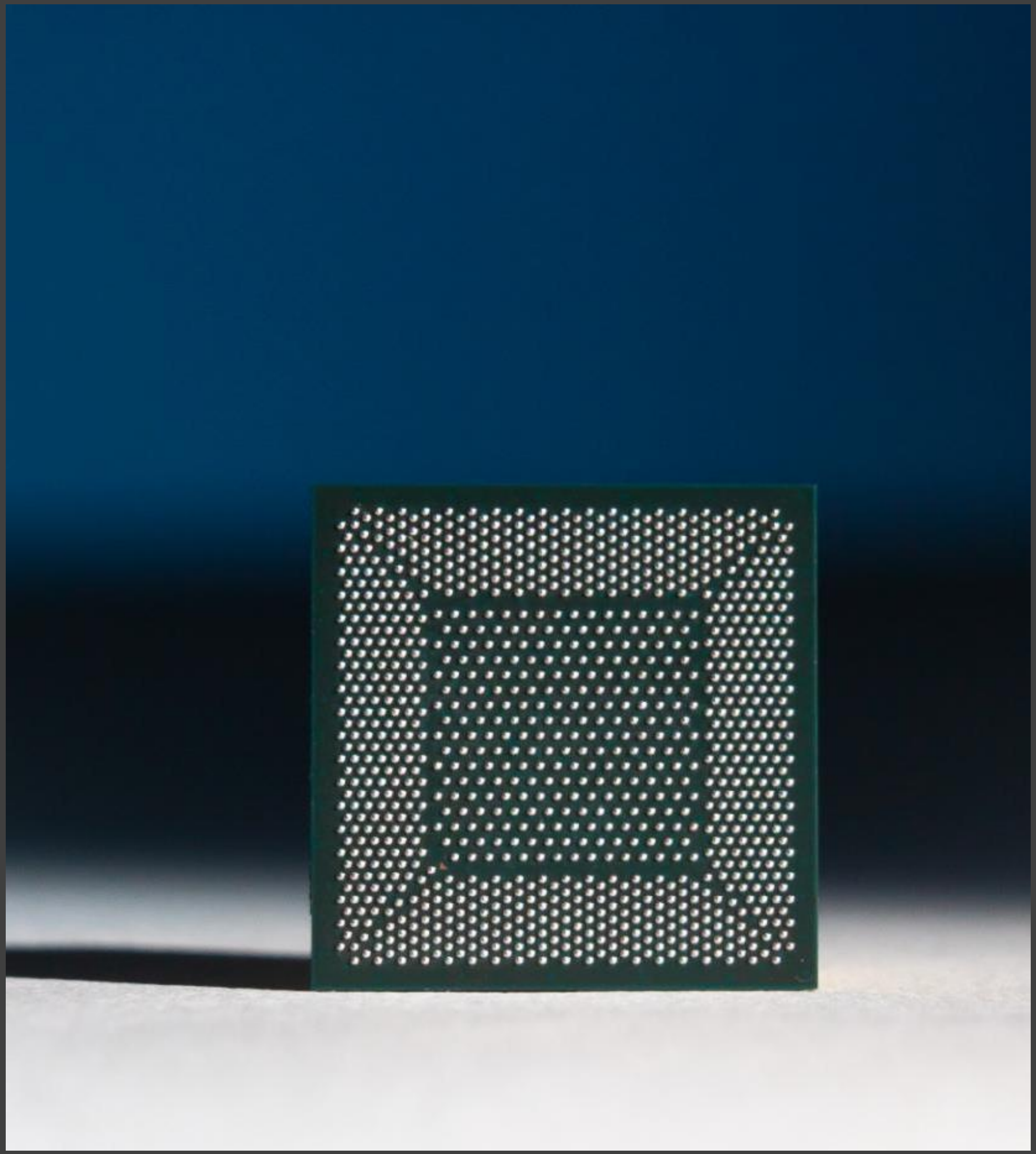
2017 : Jen-Hsun Huang (Nvidia founder): « ***Moore's law is obsolete*** »

- NVIDIA created in 1993
  - 7 G dollars of revenue in 2016-2017 with 53% only in video games (18% for data centers, 6% for car indusytry (Tesla, Toyota, Audi, Baidu, ..) and 2 G Dollars of net profit
    - From Graphical Processing units (GPU) to DNN with parallel processing (vs CPU)
    - In 2007 CUDA platform for processing any variety of DATA➔ IA, CLOUD, IOT
    - Pascal Architecture (P100) then VOLTA GPU

- OTHER GPU providers :
  - **INTEL** and NERVANA , **GOOGLE** : Deep Mind; TensorFlow (Open Source since 2015), **AMD** and its processors: ZEN; VEGA GPU
  - Cerebras, Knupath, Grapgcore,…

- « **NEUROMORPHIC COMPUTING** » (INTEL) with LOIHI research chip :
  - *emulating the neural structure and operation of the human brain,*
  - 3rd generation of AI
  - "SPIKING NN" : Each "neuron" in the SNN can fire independently of the others

https://www.intel.de/content/www/de/de/research/neuromorphic-computing.html



The Loihi research chip includes 130,000 neurons optimized for spiking neural networks (SNN) <on 14 nm>; 1 nm = 1.0E-9 m

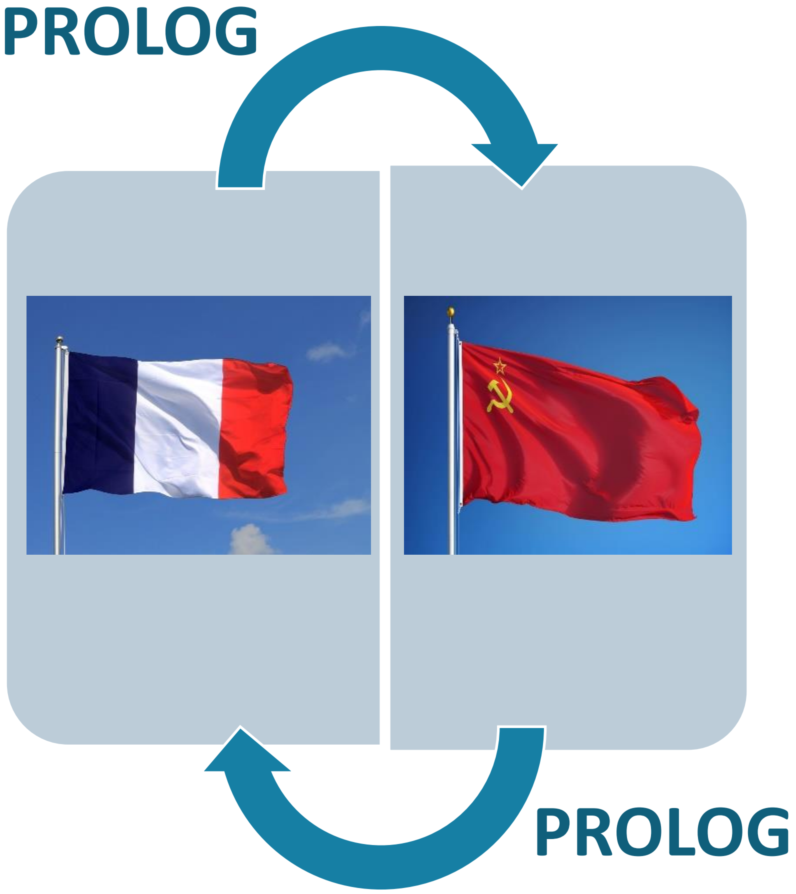
AI + BIG DATA ➔ CHATBOT (CHAT + roBOT) application replacement ?

- **Alan Turing (1950) and ELIZA !**
- Gartner :
  « 85% of Internet transactions will be performed without human beings before 2021 »
- OVUM : *« 75% of growing rate for Chatbot market between 2016 and 2021 »*
- 2016 : FACEBOOK opened its interface for Chatbots (then Youtube/Microsoft,..)
  ➔ > 10 000 Chatbots in 2 years *
  - tweets generated by robots (1/3 in the USA!)
  - Price alert , Customized press reveiew, …
- OPEN SOURCE Chatbot generator, **Chatfuel :**

http://www.leptidigital.fr/reseaux-sociaux/creer-chatbot-messenger-8755/

2 TYPES of AI . An example with PROLOG and natural language processing with Pr Alain Colmerauer in 1980

INPUT : « L'esprit est fort mais la chair est faible »
(The spirit is willing but the flesh is weak)



Double OUTPUT : **« La vodka est forte mais la viande est molle »**
(*Vodka is strong but the meat is soft*)

SYMBOLIC **AI (GOFAI)**

vs

CONNECTIONIST **AI (Deep Learning)**

2 types of AI & 2 major forms of intelligence: RATIONALE (reasoning) and INTUITION

INSTITUTE OF TECHNOLOGY

1) **Symbolic AI** < rule languages and RATIONALE>

IF THEN ELSE rules
Logic (Ex Prolog)

- **Handwritten programs**
  Ex: Expert systems such as Mycin for infectious diseases (600 rules) or Airbus for the accuracy of its airliner control software.

  Hierarchical inference engine with backward chaining
  Modeling expert knowledge

- ***GOFAI (Good old-fashioned AI) !***

- *« Instead of trying to produce a programme to simulate the adult mind why not try to produce one which simulates the child’s. If this were then subjected to an appropriate course of education (training) one would obtain the adult brain »*

Alan Turing, (Computing machinery and intelligence »
MIND, Vol 59, N° 236, Oct 1950

eBIHAR

Connectionist AI

"DEEP LEARNING is part of the future of AI" Yan Le Cun (Dec 2019)

2) CONNECTIONIST AI < Neural Networks /DEEP LEARNING/ML and **INTUITION** >

"Reasoning is only a part of human intelligence .... Perception, intuition and experience are all learned abilities and trained."

- CONNECTIONIST AI < Neural Networks /DEEP LEARNING/ML and INTUITION (antisphexity!) >

- Having a Machine CAPABLE TO LEARN
  - DATA driven (supervision, non-supervision, reinforcement)
  - Engineering by DATA
  - DATA LEARNING

- No programming but training a machine to perform a task (learn a task from examples)

- **Intuition? *"Antisphexity"*!**

- **IA tomorrow is a hybrid system with ML/DL, GOFAI and classical computing**
  - A DL system is not capable of logical reasoning.
  - Logic today is incompatible with learning
- "*The challenge for the coming years is to make them compatible*" Yan Le Cun

Biological NEURONS, brains and computers

- Biological NEURONS with DENDRITES (inputs) and AXONS (outputs)
  - <**100 billions* NEURONS** in our brain (20 in our Cortex)
  - Interconnexion among them thru SYNAPSES
    - **10 000 synapses/neuron < 1 peta synapses>**
    - In one brain cm**3 : 10 000 billions* SYNAPSES
    - Each synapse can perform a computation several hundred times per sec.
      (theoretical synaptic computing capacity of 100 PETA op. per second)
    - Average power consumption of 25 WATTS
      Information speed among synapses : 120 m/sec (430 km/h) and **1 PETA bytes** of stored data!
    - Learning: creating synapses, removing synapses or changing their effectiveness (Same logic in deep NR)

    - **On an intelligence cursor from 1 (mouse) to 100 (human), the AI D today would be 1.1**

*Billion : 10**9

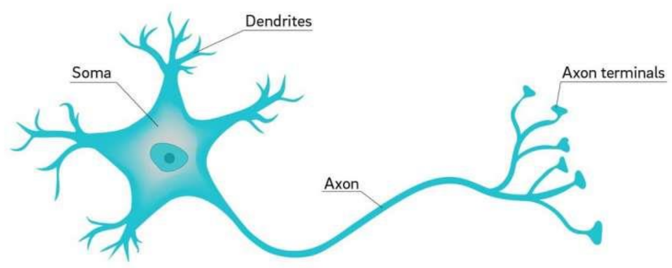
- ➔ possible simulation with a computer of **1 billion* op/sec** in 2020

(NOTE : 1 billion more powerful computers by 2050)*

Supercomputers are approaching this computing power but with more energy!
GPU card: > 10 teraflops/SEC (100 000 cards for brain computing power!)
Note: a GPU card consumes 250 WATTS (10 times more than the brain)

In 1977: The 160 Mflops Cray 1 was worth 8 million dollars (5 tons and 115 Kwatts).
GPU card 60 000 times more powerful for 300 Euros! (soon in smartphone)

Neuron



Digital neuron, perceptron and neural nets

« *Simplified view of a natural neuron which could be useful* » Alice Guyon

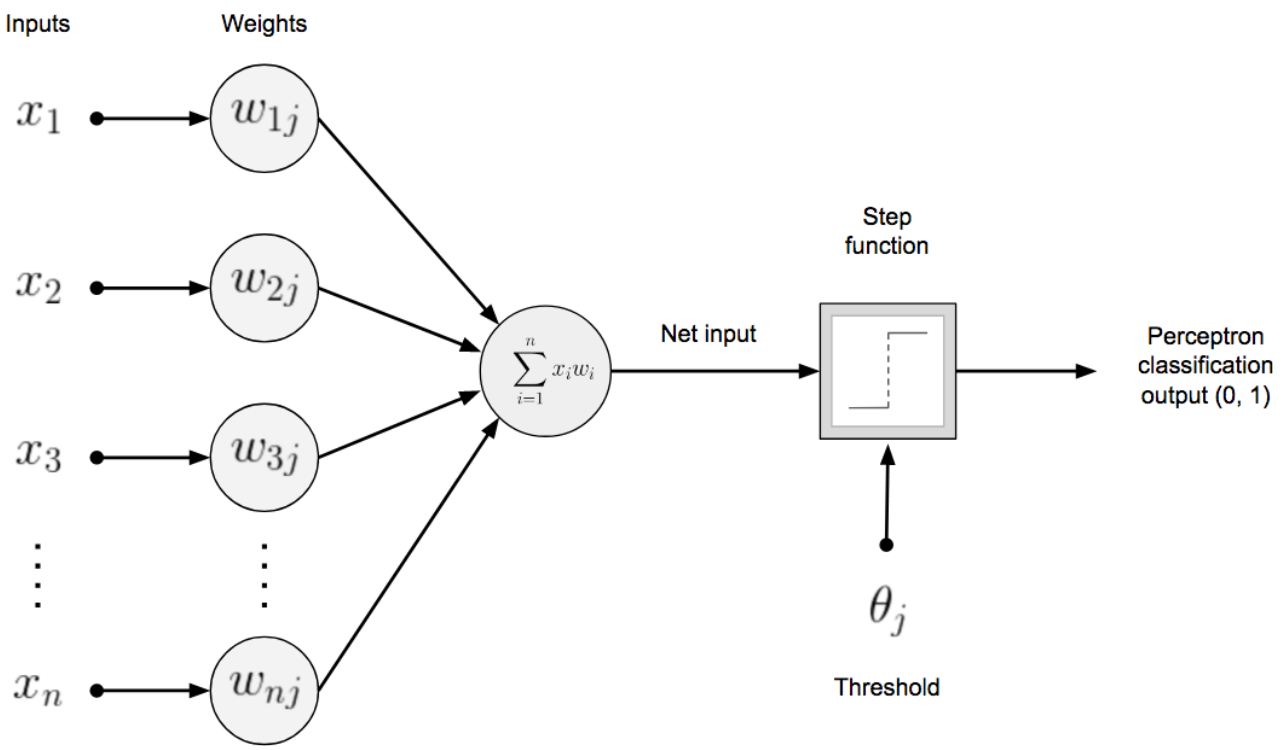


FIGURE : Typical ***PERCEPTRON*** (1957) view of one digital Neuron …

From single neuron to NN
➔ multi-layered perceptron (MLP) or NEURAL NETS (with activation function)

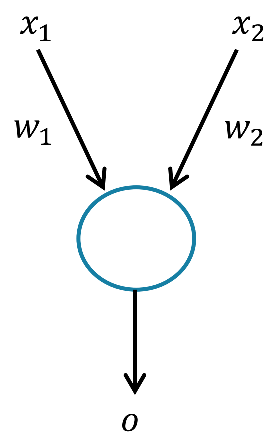
Linear algebra is the bedrock of Deep Learning :
Ax = b in basic machine learning with the matrix A , the parameter vector x , to get output column vector b

X VECTOR : {w1j, w2j, ..wnj}

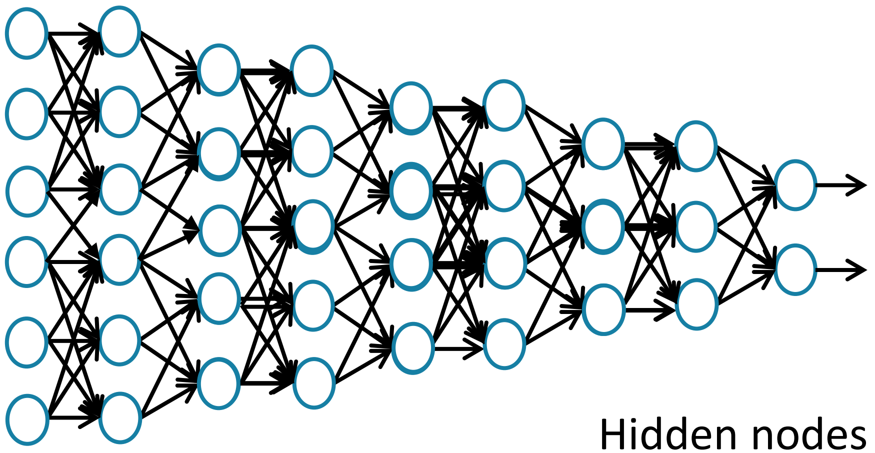
Deep-Learning Neural Network (DNN)

- DEEP LEARNING is a rebranding of **NEURAL NETWORKS (<u>with more than 2 layers</u>)**
- DNN is a **multi- layer** Neural Net which can be processed efficiently with **GPUs** (*Graphical Processing Units*) and trained with Big Data (**data-centrics learning**)
- Millions of Neurons in HIDDEN LAYERS with
  - no intra communication among neurons within each layer

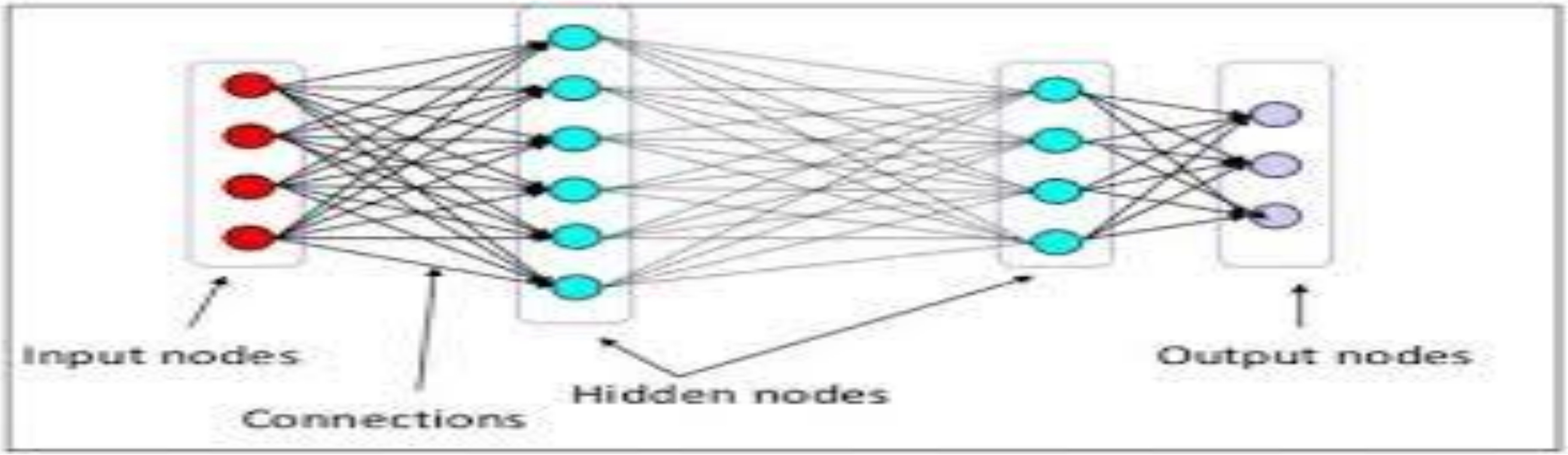
- **Neurons**
- Computational units
- **Connections**

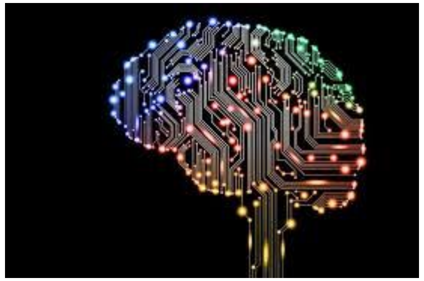


- **Weight** inputs from previous layers before feeding into next layer

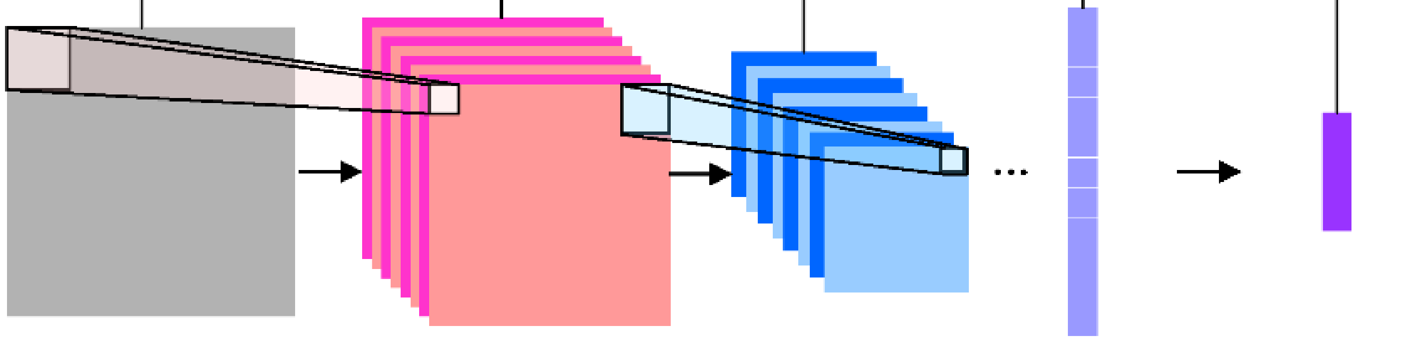


Deep learning



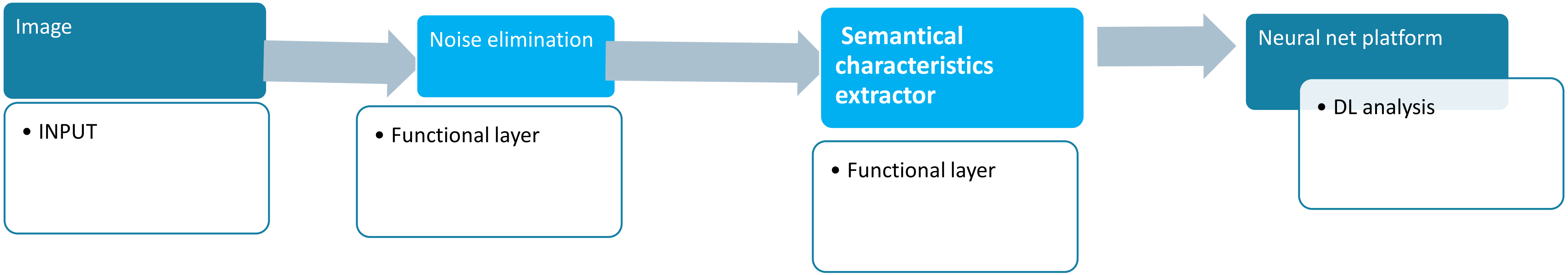


Convolutional neural nets (CNN or CONVnets)



- **Convolutional layer (CONV)** – The convolution layer (CONV) uses filters that perform convolution operations. Its hyperparameters include the filter size *F* and stride (step) *S*. The output is called *feature map* or *activation map*.
- **Pooling (POOL)** – The pooling layer (POOL) is a down-sampling operation to reduce size of an image without loss of important information.

Pipe Line architecture & DEEP LEARNING



Data preparation and noise elimination (Ockham’s razor)

- « *Everything shoud be made as simple as possible but not simpler* » Einstein

Last functional layer before DL platform : « **characteristics extractor** »

- In manual mode :
  - Ex : SVM (*Support Vector Machine*) method from Isabelle Guyon (Orsay) well formalized but … « ***far from reality*** »(Yann Le Cun) !
- **In (data) learning mode** : « ***DL challenge*** » (Yan Le Cun)

AI "is a diligent assistant who works fast without fear of repetitive tasks".

- *"INTUITION for me has always preceded mathematical FORMALIZATION ...avoiding the evil of the French search for mathematical beauty per se (Ex SVM of the Univ d Orsay or DATA LOG)"* Yan Le Can

- Example : Multilayer Neuron Networks cf CAT vision (2 Nobel awards in Biology)
  - 1981 : **two neurobiologists (D.Hubel and T.Wiezel)** are awarded the Nobel Prize for their work on the cat's VISUAL system composed of several layers of neurons :

    retina to the primary visual cortex (simple cells) then to the inferior temporal cortex (complex cells)

    ➔ NEOCOGNITRON machine of Fukushima (Japan)

    ➔ CNN (Convolutional NN)

- Artificial neuron with neuron = **Simple mathematical function** with input/output vector and intermediate matrix

Some other disruptive technologies in the data economy

- NFC (Near Field Communication)
- LIFI (Light Fidelity)
- **Blockchain**

NFC and… MOBIQUITOUS systems ?

Tweet on Mobiquity*

Serge Miranda
@SergeM06

MOBIQUITY ? = MOBIlity of the cell phone who became a computer (a "smartphone") and ubiQUITY of Internet who became social and broadband

Reply Delete Favorite

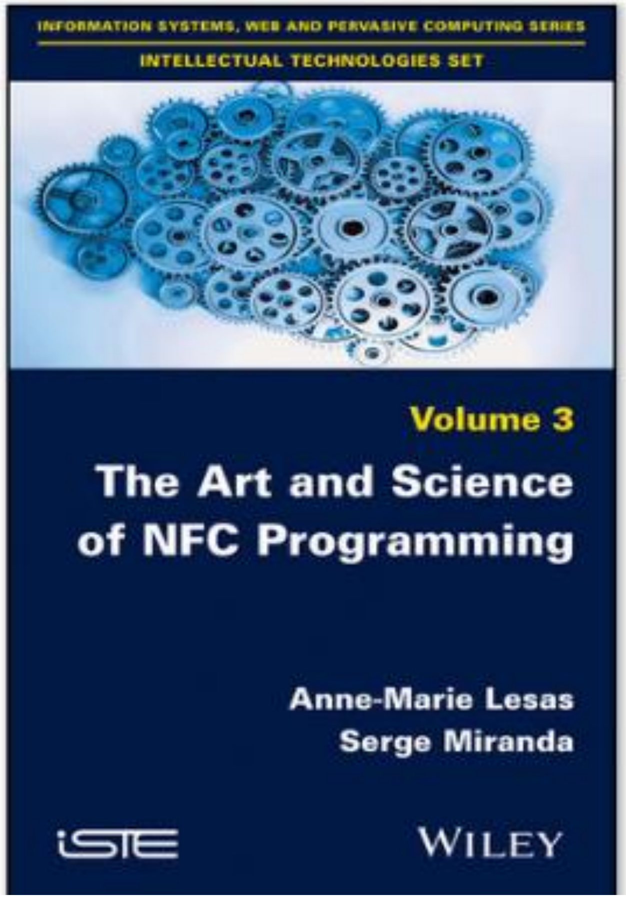
* *Part of a « TWEETED SEMINAR » on NFC Technology and Mobiquitous information systems at University of Tampa, Florida (2012) < Tweet Presentation on Slide Share>*

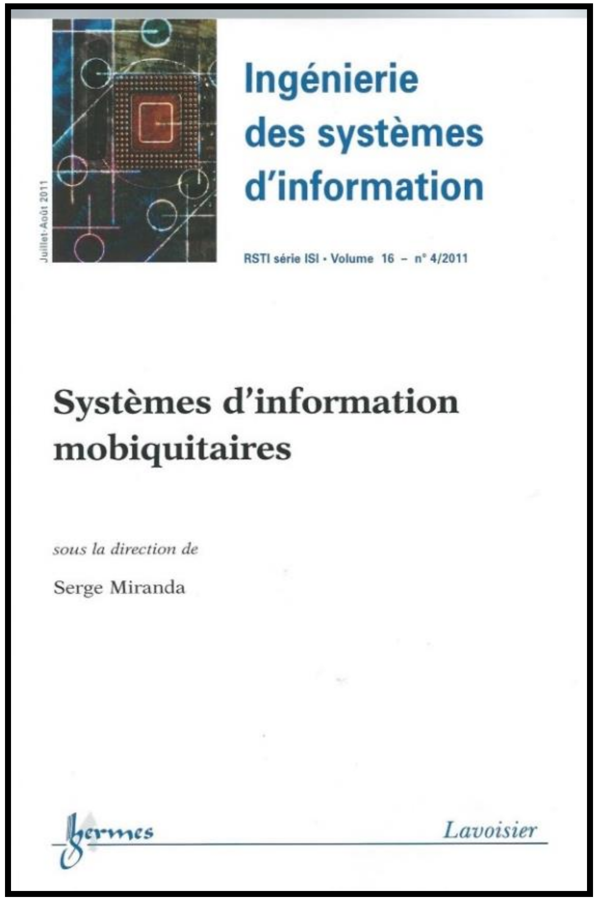
ESTIA INSTITUTE OF TECHNOLOGY

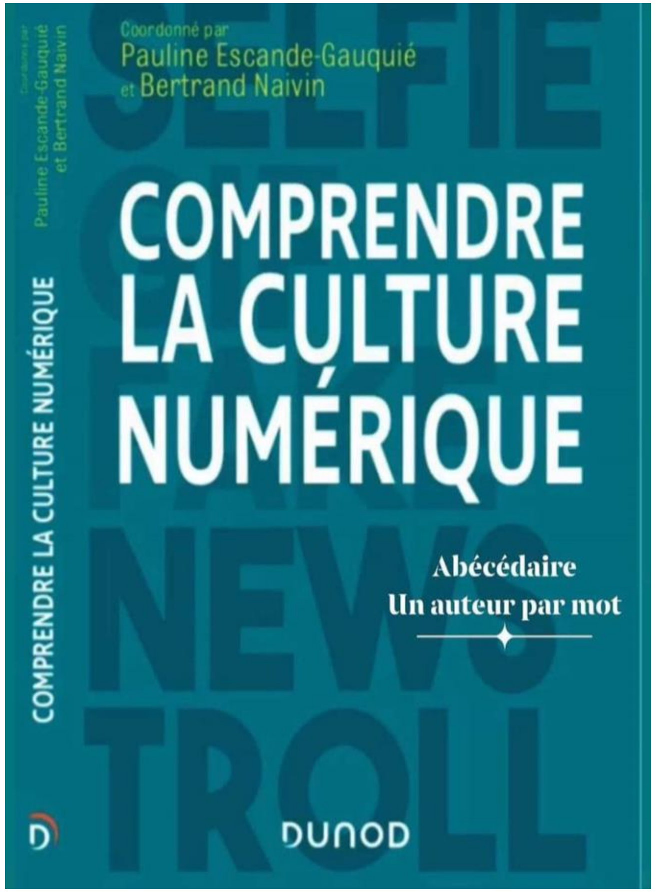
e-BIMAR

MBDS books on NFC Programming and mobiquitous systems (2011 & 2017 & 2019)

INSTITUTE OF TECHNOLOGY







e-BIMAR

« Homo Mobiquitus & commonactors »

50% of the planet owns a smartphone in 2015 (50% being NFC)

INSTITUTE OF TECHNOLOGY

Bouthan

Kirghistan

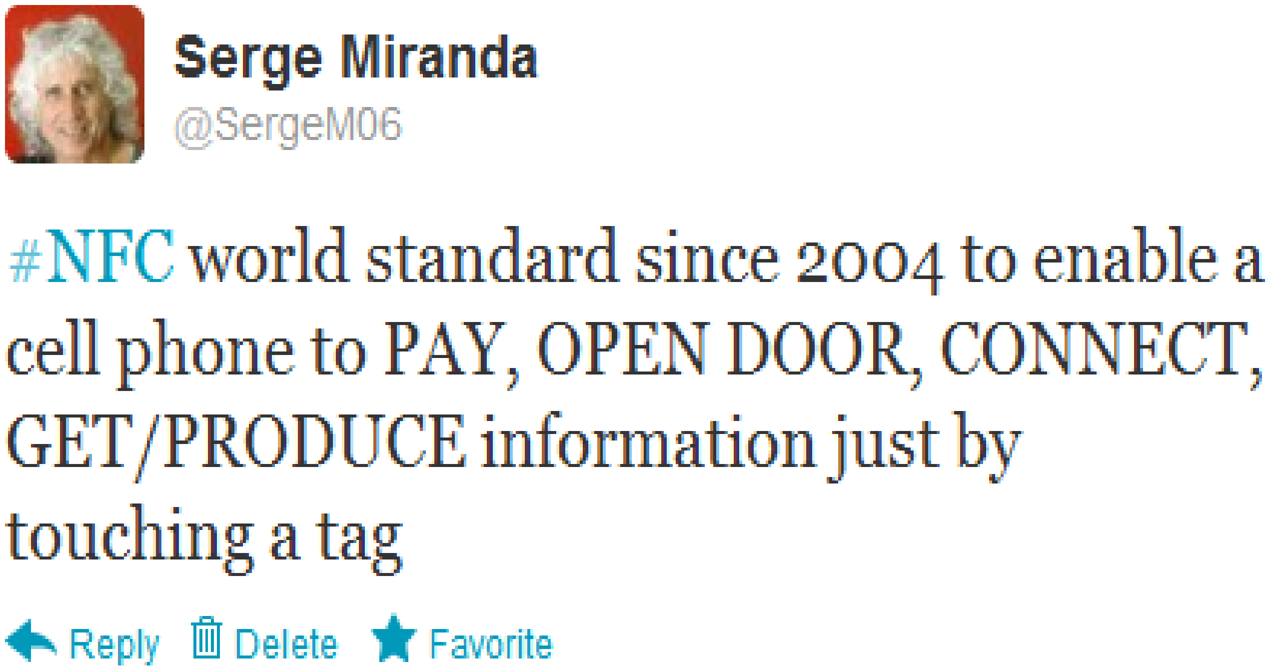
Sister Flora
(Haiti)

e-BIHAR

NFC (world) STANDARD

INSTITUTE OF TECHNOLOGY

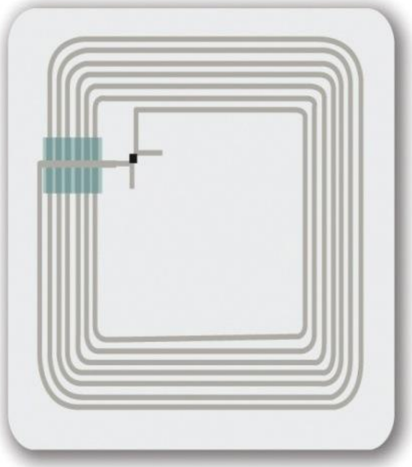
NFC
(Near Field Communication) :
« TOUCH'n PLAY » for universal connector…



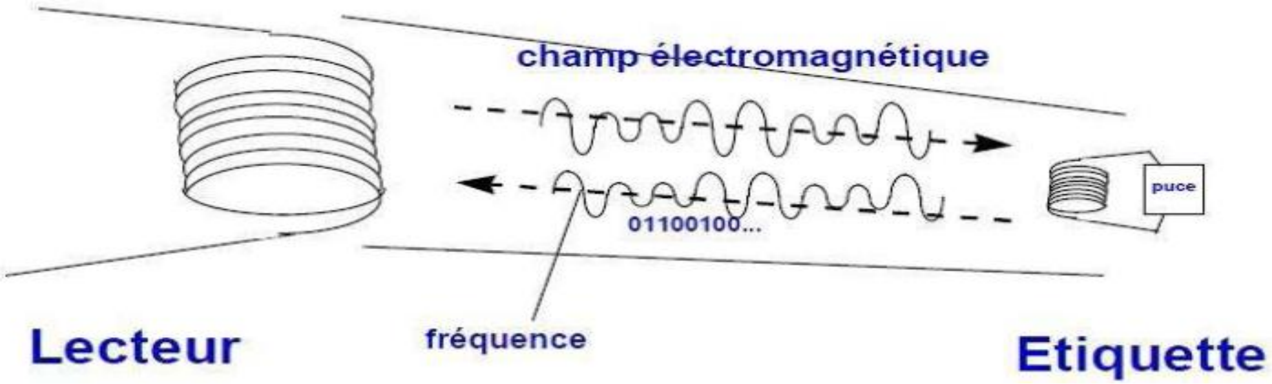
(TWEET Seminar at University of TAMPA in 2012)

eBIMAR

NFC/ RFID
(Radio Frequence IDentification) tag ?



RFID tag from ASK
(Sophia Antipolis)



The three modes of NFC standard

1. **« READ/WRITE »** (active/passive)
2. **« CARD EMULATION »** (passive/active)

   « SECURE ELEMENT » (SIM or ?) or HCE (Cloud)

   TSM (Trusted Service Manager)/SE
3. **« PEER TO PEER » P2P** / (Active/active)

The five « W » of NFC applications

2) WHERE &
3) WHEN ?
(Ici et maintenant!)

4) WHERABOUT ?
(GOAL)

5) WHAT ?

1) WHO ?
(USER Profile)

« Mobiquitous effect »?
(combination of Moore's, Morris', Metcalfe's, Gilder's laws)

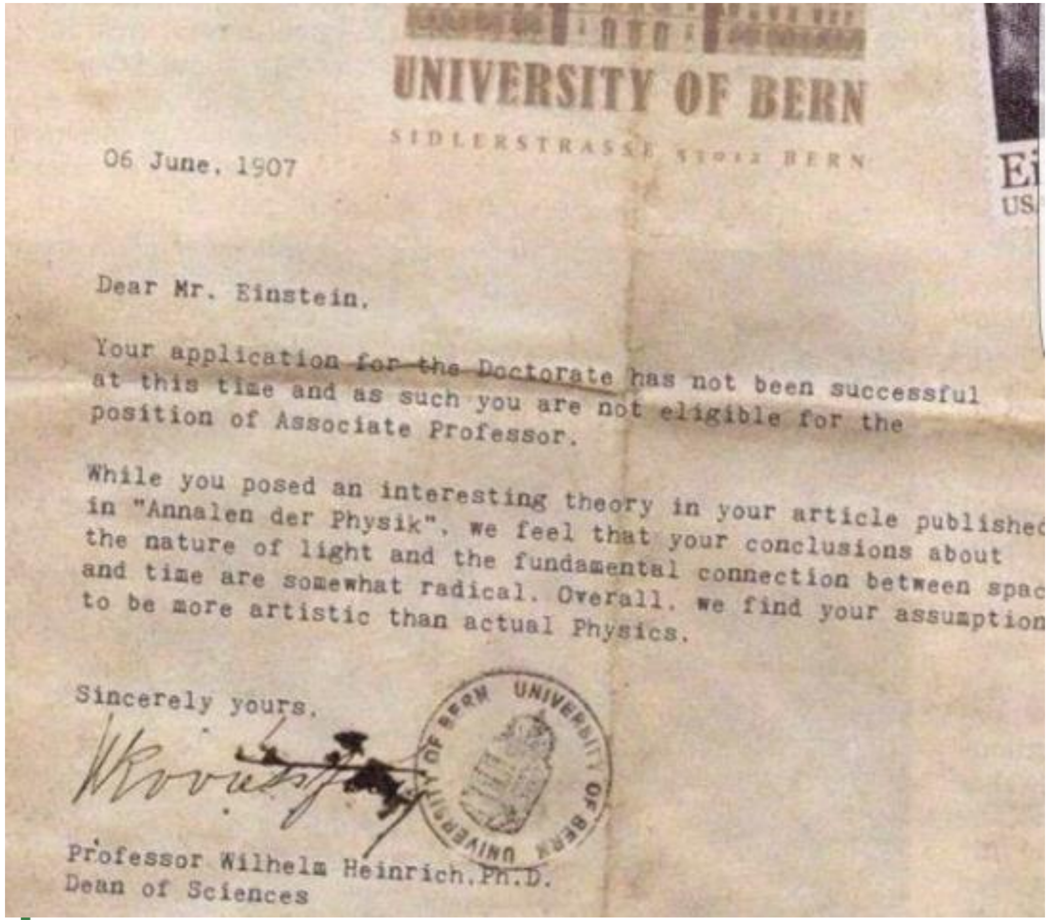
- **Gilder's law on bandwidth**
  - *Bandwidth doubles every year and half*
- **Metcalfe's law on Network value**
  - *Network value depends on squaring the number of its nodes*
- **Morris law on storage**.
  - *Number of gigabits per square inch doubles every year*
- **OC law on smart/tagged objects (IOT)** :
  - *Number of smart objects doubles every year*
- **(BIG) DATA law**
  - *The quantity of data doubles every year*

- **Moore's law on computer power :**

The CPU processing power (number of transistors per chip) doubles every 18 months (100 times per decade)

until 2030 then obsolete and… new NVIDIA- founder's law on GPU !

1st DATA equation for the future: E=MC2



UNIVERSITY OF BERN

SIDLERSTRASSE 5 3012 BERN

06 June, 1907

Dear Mr. Einstein,

Your application for the Doctorate has not been successful at this time and as such you are not eligible for the position of Associate Professor.

While you posed an interesting theory in your article publishe in "Annalen der Physik", we feel that your conclusions about the nature of light and the fundamental connection between spac and time are somewhat radical. Overall, we find your assumption to be more artistic than actual Physics.

Sincerely yours,

Professor Wilhelm Heinrich, Ph.D.
Dean of Sciences

- **E = MC2 !**

Energy
Multimedia
Computer &
Communication

2nd DATA equation for the future : E= MC3 !

« E= MC3 »

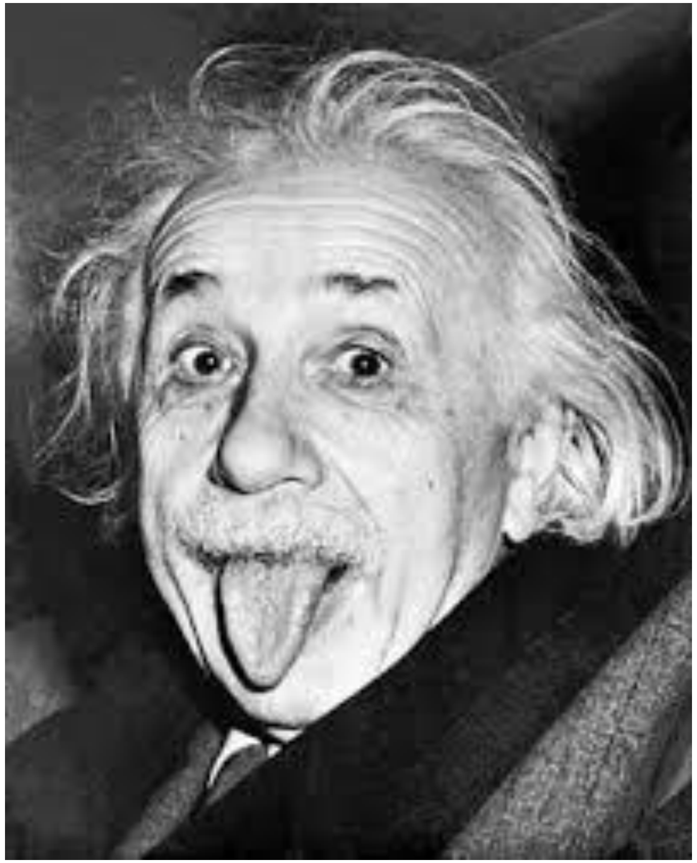
Energetic resource

M obiquitous systems

Computer **+**

Communication **+**

Consumer Electronics



LIFI (LIGHT FIDELITY)

Harald HAAS

- **LIFI in 60 seconds <in French>**
- https://youtu.be/RtMmKBQJz6k

LIFI (Light Fidelity) ?

- **HIFI (High Fidelity) then :**
  - **WIFI** (RADIO portion of the electromagnetic spectrum)
  - **LIFI** (LIGHT sprectrum from LEDS)
    - communication technology based on the LIGHT spectrum and the use of the VISIBLE BLUE (450 nm) and THE VISIBLE RED (760 nm) of lights generated by LEDs
    - **LED includes a semiconductor :**

      Fast switching of light (LEDs are semiconductors : 10**9 per second !)
- **Long life + LOW POWER + HIGH DATA RATE** (10 times WIFI) + **SAFE** ➔ HD Streaming
- **Standardized : IEEE 802.15.7 since 2011**

LIFI applications (with location-based services)

INSTITUTE OF TECHNOLOGY

- OUTDOOR applications in smart cities :

Smart Public Lighting

- INDOOR applications
  - **Smart Homes**
  - Smart Offices
  - Smart Campus
  - Smart Airports
  - **Smart Hospitals**
- **Smart MUSEUM**

LIFI demo video (in French) of Mbds JMAGINE* project in the International Museum of Parfume in Grasse (France) :

https://www.youtube.com/watch?v=tHKG9CkQm7U

**Open Source platform for teachers to create invisible paths in smart territories (and smart museum)*

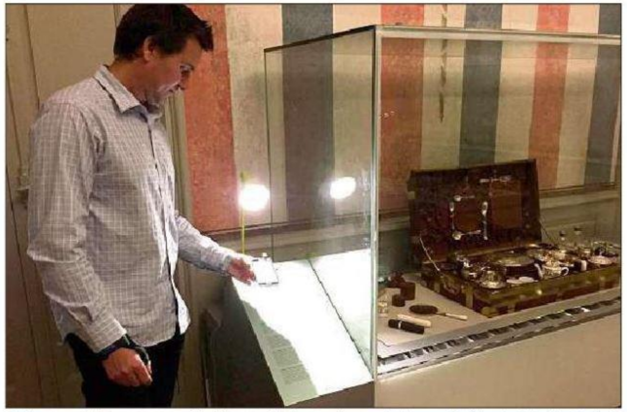
Le LiFi, WiFi du futur, décrypté au MIP

Et s'il existait un outil de communication encore plus moderne que le WiFi?
Une conférence au MIP, mardi soir, permettait de faire toute la lumière sur le LiFi : une technologie de communication sans fil utilisant l'oscillation de lumières LED.
Le principe est simple : il suffit de passer son smartphone équipé d'un Dongle ou la caméra du smartphone – s'il est compatible – sous un faisceau lumineux pour accéder à du contenu.

75 % de LED en 2020

Au Musée International de la Parfumerie, trois prototypes avaient été mis en place. En passant son téléphone sous la première lampe LED, on découvrait – en multilingue – le contenu de la mallette de Marie-Antoinette exposée dans les salles du musée. Le second prototype permettait d'accéder à une vidéo réalisée par les collégiens de Saint-Hilaire *(lire par ailleurs)* et le troisième prototype était un diffuseur de parfum très amélioré! *« Quand la lumière déclenche une odeur, c'est le début d'une créativité sans limite »*, assurait Serge Miranda.
« Notre fantasme, c'est de faire en sorte que la région devienne la capitale mondiale de la LiFi, ajoutait Marc Lassus, précurseur il y a quelques années, dans le domaine de la carte à puce. *La France est en avance ! »*
Un cocorico bienvenu.
« D'ici 2020, 75 % des ampoules seront des LED », notait encore Serge Miranda, avant de citer l'exemple de la Suède qui a banni le WiFi de tous ses hôpitaux au profit du LiFi, ou encore les militaires qui ont fait le même choix.
« La créativité sur les usages est infinie », insistait Serge Miranda. Le fondateur et directeur du master Mobiquité, Big Data et intégration de systèmes (MBDS) reçoit régulièrement des stagiaires qui viennent se former à la manipulation du LiFi. En ce moment, deux étudiantes marocaines réfléchissent à un projet d'éclairage public qui permettrait aussi de recharger son téléphone grâce à la LiFi… un projet qu'elles espèrent développer à Casablanca, la ville qui accueillera la COP22.

L. S.



Grâce au LED, on pouvait découvrir le contenu de la mallette de Marie-Antoinette... Dans toutes les langues ! (Photo L. S.)

Le collége Saint-Hilaire et la LiFi

Sous l'impulsion de Mélanie Fillon-Robin, professeur d'arts plastiques, les élèves de 4e4 au collège Saint-Hilaire ont réalisé un court-métrage de trois minutes. En utilisant la technologie LiFi, la vidéo *« s'infiltre »* dans le téléphone du visiteur et le captive.

Les collégiens se sont inspirés de *Jardin d'Addiction*, sculpture de Christophe Berdaguer et Marie Pejus exposée au MIP. Vous pouvez découvrir leur travail en *« flashant »* le QR code ci-contre via l'application dédiée sur votre smartphone.

Questions à Serge Miranda

Fondateur et directeur du master MBDS à l'université de Nice-Sophia Antipolis.

Les avantages de la LiFi ?
Il n'y a pas d'ondes électromagnétiques. La vitesse de transmission des informations est 10 à 100 fois supérieure au WiFi.

Et les inconvénients ?
Il faut être sous le faisceau lumineux. Mais le LiFi peut fonctionner même quand on ne voit pas la lumière, avec des infrarouges par exemple. C'est ciblé. C'est aussi un avantage : la géolocalisation est non intrusive.

C'est-à-dire ?
On ne consulte les informations que si on est intéressé en plaçant le mobile sous le faisceau. Si on ne veut pas être embêté, on garde son téléphone dans sa poche.

eBIMAR

Blockchain (*distributed data base*) and digital record-keeping (transaction tracking, …) & everlasting application !

- **In proof we trust**
- **« The GOD protocol »**
  Nick SZABO cf Lederman's « The GOD particle »
- **« the GOD DATA » !**
  - End of serendipity ?
- **« You are YOUR projects »** (Tom Peters)

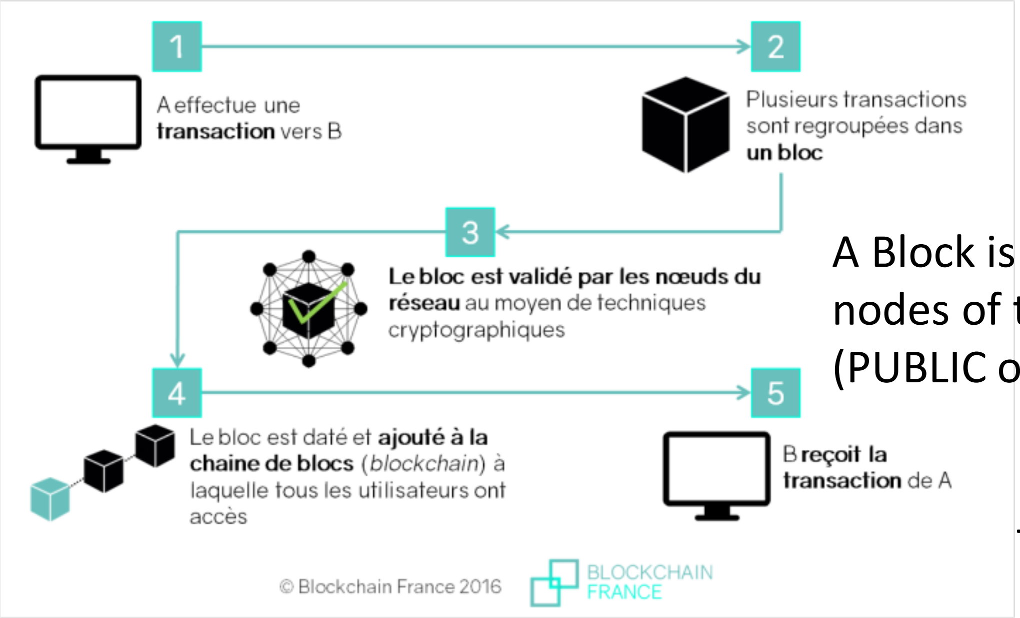
➔ « *You are YOUR DATA* » !

Blockchain technology will revolutionise far more than money :
it will change your life !

Blockchain and fully-duplicated ledger

A does a TRANSACTION to B

Various transactions are gathered into **BLOCKS**



A Block is validated by (all) nodes of the network (PUBLIC or PRIVATE)

A block is timestamped and appended to a **BLOCK CHAIN**

B receives the transaction from A

https://www.cigionline.org/multimedia/what-blockchain

DATA future (*Let us decode it !*) ?

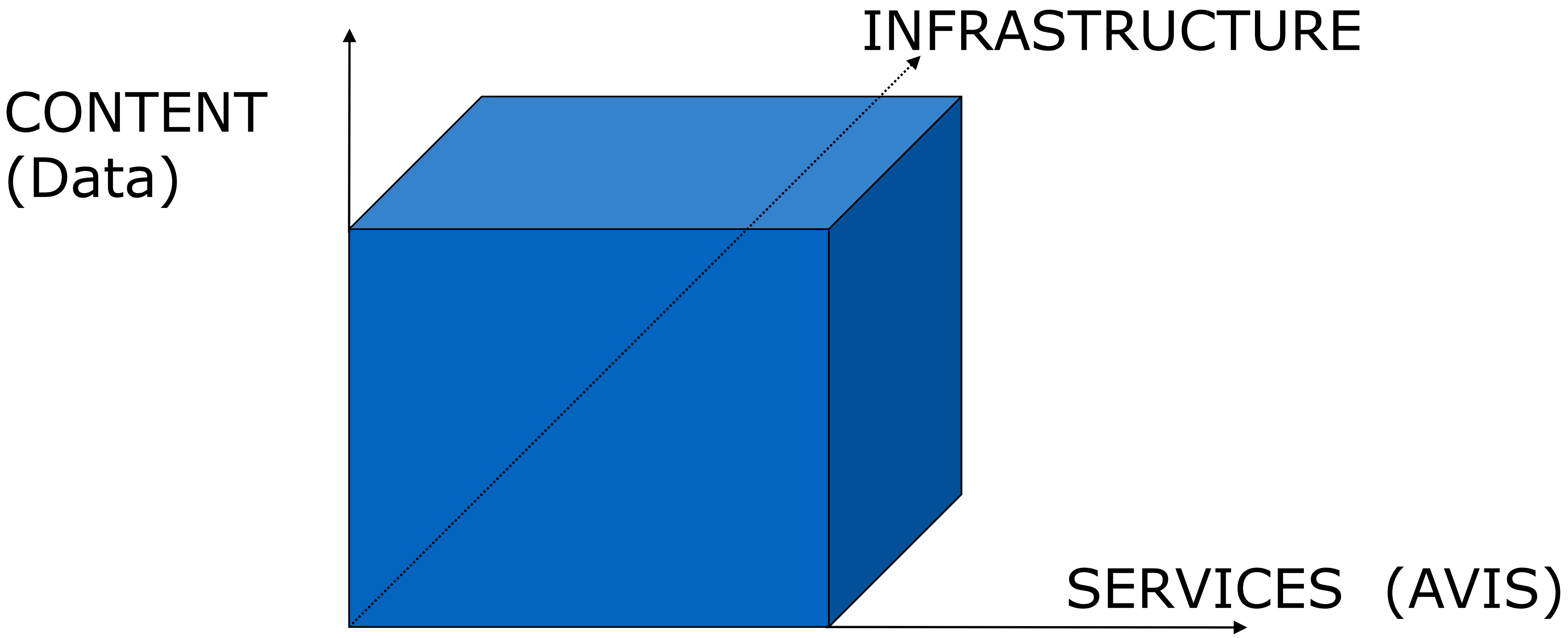
Three dimensions of the future in the DATA ECONOMY

- Three dimensions of our DATA future
  - Internet of Everything ➔ **SMART PLACES & Little BIG DATA**
  - **BOTTOM UP paradigm** (innovation, energy, computing,..)
  - **Homo Mobiquitus** and « *commonactors* »

Caricature PAVO (Conf Cnameo Toulouse, 30/9/2020)

ICT CUBE

« *A cube is a metaphor for a strong relation* » J.Olsson



From TOOLS to SERVICES (AVIS) and SMART PLACES

- « *If we can predict the future of the infrastructure we cannot predict the future of services… services cannot be controlled TOP DOWN… Digital divide on services not in technology…* **from services to SMART PLACES…** »

Pr. Leonard Kleinrock (UCLA), June 2008, Brussels

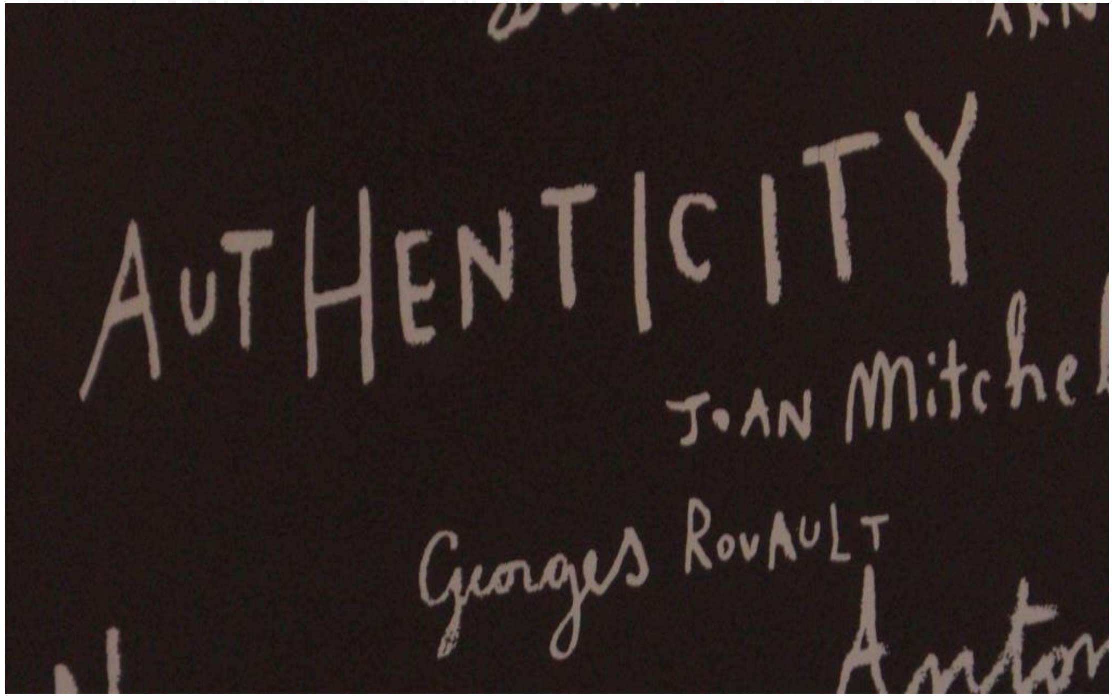
- **AVIS >> HERTZ**
  - AVIS :
    Added-value information services
  - HERTZ :
    Heoric Executive Retreat to Zero

From Services to SMART PLACES

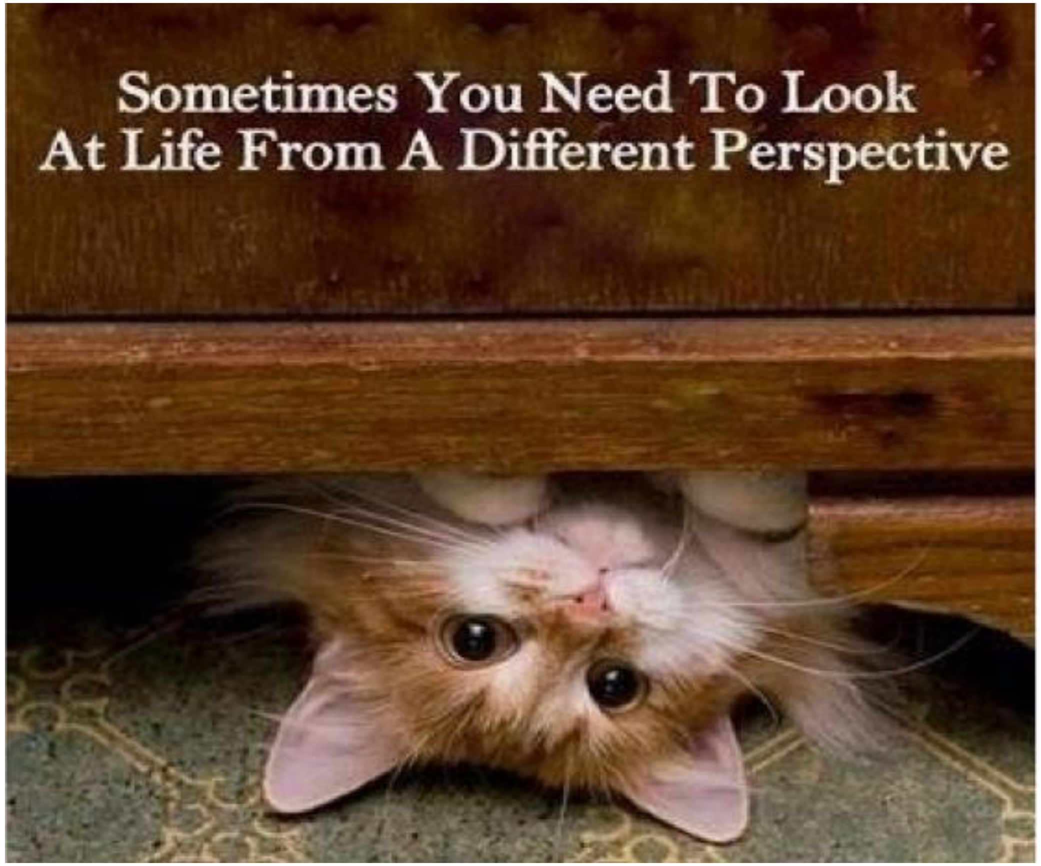
- **TOOLS ➔** *« Quantity »,*
- **SERVICES ➔** *« Quality »,*
- **SMART PLACES**

  ➔ *« Authenticity* »*

- *« *Authenticity » J.H.Gilmore, B.J.Pine, Harvard Business Review, 2007*



From TOP DOWN (infrastructure) to BOTTOM UP approach (services)



- From hierarchical
  Top down
  (1:N) of the past to the
  « Bottom Up »
  (N:1) of the future

BOTTOM UP :

1. Computer science
   (PC ! Internet, Smartphone !)
2. Renewable energy
3. Society (*Holocraty*)
4. Innovation,
5. Deep Learning, …

Homo Mobiquitus ?

- Homo Habilis
- Homo Sapiens
- **Homo Mobiquitus** (and COMMONactors*)

- ***The Smartphone won the battle of the pocket*** !

*Serge Miranda, « *New data territories/Nouveaux Territoires numériques* », Book, Ecole des Mines, Nov 14 [MIRA2014]

Society evolution

- *CONSUMPTION* (**production**) society
- ***COMMUNICATION*** society
- ***COMMONACTION*** society

- ***« The best way o find yourself is to lose yourself in the service to others »*** GANDHI

Society evolution ?

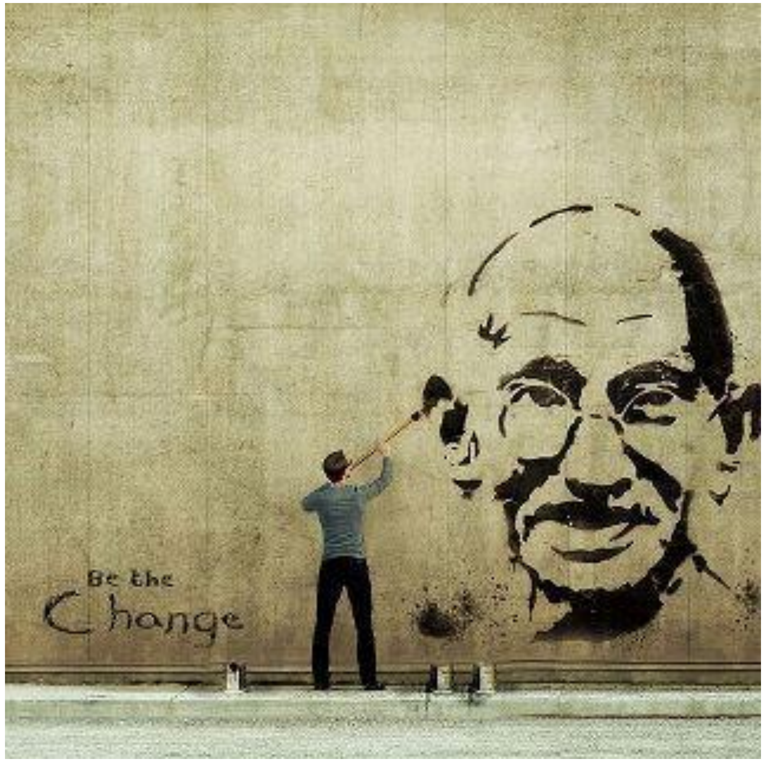
- « *Society life in which reigns the* ***PRODUCTION Mode*** *will come out with a large accumulation of GOODS* » Karl Marx – The Capital (1867)
  - ➔ **Industrial revolution**

- « *…(**COMMUNICATION mode**)… a large accumulation of SHOWS* »
  Guy Debord (1967)
  - ➔ **Internet revolution**

- ***« …RECOMMENDATION/COMMONACTION mode …a large accumulation of DATA »***
  (2017)
  - ➔ **DATA revolution**

« Be the change you want to see in the world»

GANDHI



eBIHAR

Towards a student head …

- Well « **FULL**»
  with Writing
- well « **MADE** » (Montaigne)
  with Printing
- well « **CONNECTED** » (Michel Serres)
  with Internet

- towards well « **AUGMENTED** »
  with mobiquity/Smartphones,
  AI and Little Big Data !

DANGER : « ANYWHERE » vs « SOMEWHERE » *

INSTITUTE OF TECHNOLOGY

- Digital divide on
  - Social CULTURAL aspects
    (identity, Ecology, Immigration, flexibility) vs
  - Social ECONOMICS
    (Consensus on regulated free trade with social and education framework)

- ANYWHERE :
  flexible urban persons of the Knowledge data society with GLOBAL trade and international living
  vs SOMEWHERE (deeply rooted with LOCAL living)

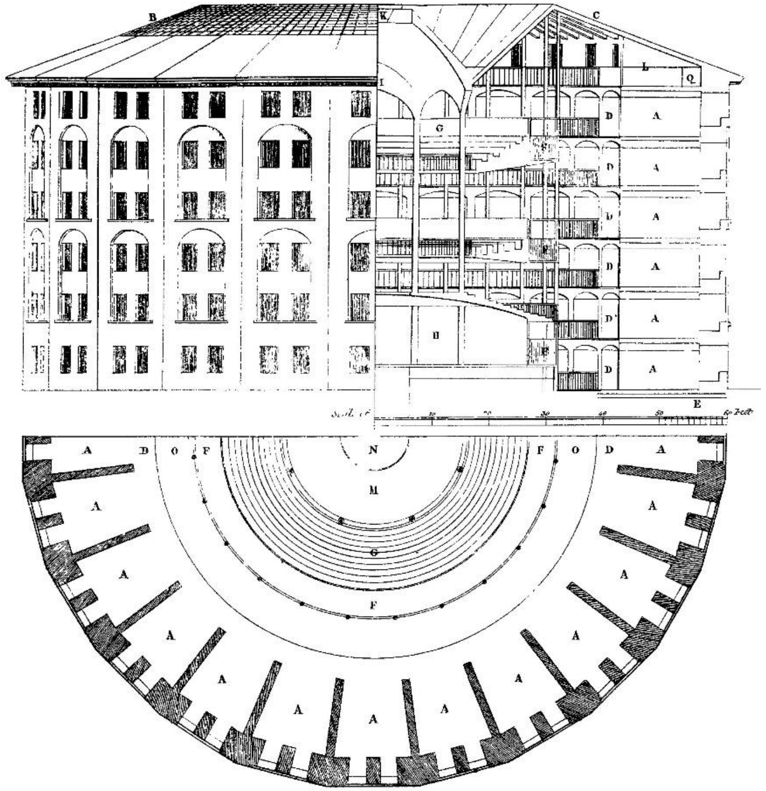
DANGER (domination) : ANYWHERE>> SOMEWHER

* *David Goodhart, « The road to somewhere », 2017*

Danger : « Panopticon* 2.0 » (*CONFINEMENT 2.0*)

- Less freedom for more security !?
  - *Everybody can KNOW anything on anybody*

- Jail architecture with a factory model imagined by **Jeremy BENTHAM** to enable a centralized guard to view and control every prisoner*



* ***(Panopticon by Bentham, philosopher and architect, 1780)***

Conclusion : *Spiralist* Innovation

INSTITUTE OF TECHNOLOGY

In the data economy :

« **EVERYTHING is SPIRAL** »

« SPIRAL is aesthetics of CHAOS »
« Spiralism is LIFE… Everything is SPIRAL ! »
Franketienne*, *2012*

**Creator of SPIRALISM concept in litterature*

DATA-centrics spiralist Innovation ?

- Innovation ?
  - INVENTION meeting an USAGE !
  - **Bottom up and multidisciplinary**
  - (traditional academic research is top down and mono disciplinary)

- Quadrants of digital innovation by Pr. Gary PISANO* (Harvard)
  - ROUTINE Innovation
  - TWO DISRUPTIVE INNOVATIONS :
    - **Disruptive TECHNICAL Innovation**
    - Disruptive ECONOMICS (Busines-model) Innovation
  - ARCHITECTURE Innovation

  - * Gary Pisano « *You need an innovation strategy* » Harvard Busines Review, June 2015 (in French Gary Pisano, Harvard Business Review, Summer 2016, pp 16-25)

Innovation quadrants by Pr. Gary Pisano

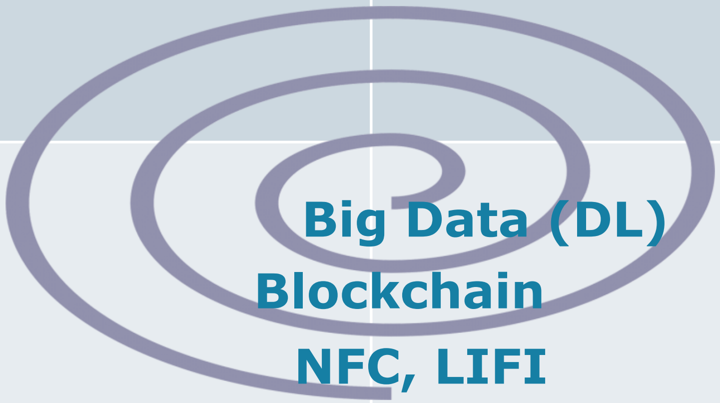
	Old technologies	New Technologies
NEW BUSINESS MODEL	**Disruptive economics (BM) Innovation** *Sharing Economy (Uber, AirB&B, ..)*	**Disruptive ARCHITECTURAL INNOVATION** *APP STORE* *CLOUD*
OLD Business model	**ROUTINE INNOVATION** *New car, new smartphone*	**Disruptive TECHNOLOGICAL INNOVATION** *BIG DATA, NFC, LIFI, Blockchain, Deep Learning*

SPIRALIST INNOVATION (some kinetics)

	OLD TECHNOLOGIES	NEW TECHNOLOGIES
NEW BUSINESS MODEL	**Disruptive economics INNOVATION**	**Disruptive architectural INNOVATION**
OLD BUSINESS MODEL	**Routine INNOVATION**	**Disruptive technological INNOVATION**

SPIRALIST INNOVATION in the DATA ECONOMY

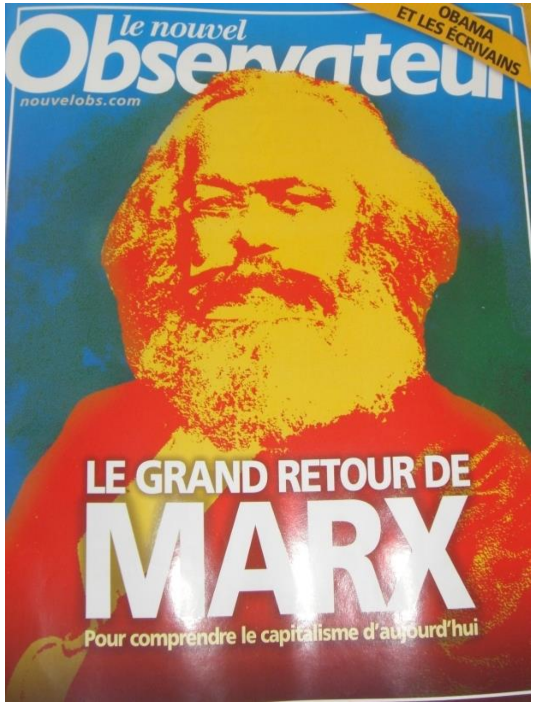
	OLD TECHNOLOGY KNOW-HOW	NEW TECHNOLOGY KNOW-HOW
New Business Model (BM)	**disruptive BM innovation**	**ARCHITECTURE INNOVATION**
Old Business model	**ROUTINE INNOVATION**	**TECHNICAL DISRUPTIVE INNOVATION**



CONCLUSION 1
Sister FLORA (Haïti) & Karl Marx !

«If you cannont change the world try to change YOUR world »

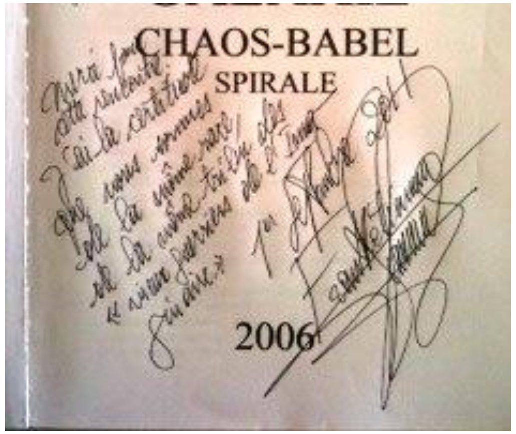
Marx
(last sentence)



« ***An ant can bear an elephant*** » *Sister* Flora (Haiti, June 2013)

CONCLUSION2 : Be « *innovation warriors* » !

INSTITUTE OF TECHNOLOGY



« Tribe of IMAGINATION warriors »

« J'ai la certitude que nous sommes ..
de la même race,
de la même tribu des
VIEUX GUERRIERS DE L'IMAGINAIRE »

FRANKETIENNE (1/9/2011)

eBIMAR

Questions ?

< DATA humanum es ☺> <The world as a neural net (V.Vanchurin 2020> < Category theory as a uniform formal data model for polystores>

« *If you are crazy, it is possible. Remain open to creativity and innovation* » John Gage, 28 Oct 99 MBDS Sophia

« *Stay hungry, stay foolish* » Steve Jobs, Stanford, 2007

Charles Babbage en 1812 :

« Propose to any english man any principle or any new instrument however admirable and you 'll observe he will spend his energy to demonstrate it could 'nt work.

Propose it to an american, he will congratulate you and spend his energy to find new applications »

e-BIHAR

The two first European GRADEOs in January the 4th with three complementary concomitances

INSTITUTE OF TECHNOLOGY

(i) **eBihar* MSc launching** on FUN platform on January the 4th (one video-tutoring session from January to May with a genuine Oracle Learning Subscription) for initial education

https://www.datumacademy.com/masters

(ii) **BIHAR/ESTIA GRADEO (microcredential) on « BIG DATA and AI » starting** on January the 4th (the 2 first GRADEOSin Europe) for continuous education; registration on FUN MOOC platform:

- GRADEO 1 :https://gradeo.fun-mooc.fr/sql-programming
- GRADEO 2 : https://gradeo.fun-mooc.fr/big-data

(iii) **New Oracle University Learning Subscription** starting in January with key data (cloud) expertise to provide professional complementary courses within BIHAR Gradeos

Succesful launching of GRADEOS with FUN, ORACLE, University of Siena, ESTIA and Datum Academy on December the 2nd 2020 (300 on-line attending people))

OTHERS …7 free webinars on AI and BIG DATA in December 2020, January 2021 and February 2021

NOTE : * ***BIHAR*** *(**B**igdata **I**ntelligence for **H**uman **A**ugmented **R**eality) also means « **TOMORROW** » in Basque language*

eBIHAR

Yet to come in 2021 around Bihar MSc

- 6 webinars on AI and Big Data, 7 GRADEOs in English (January) and in French (October)

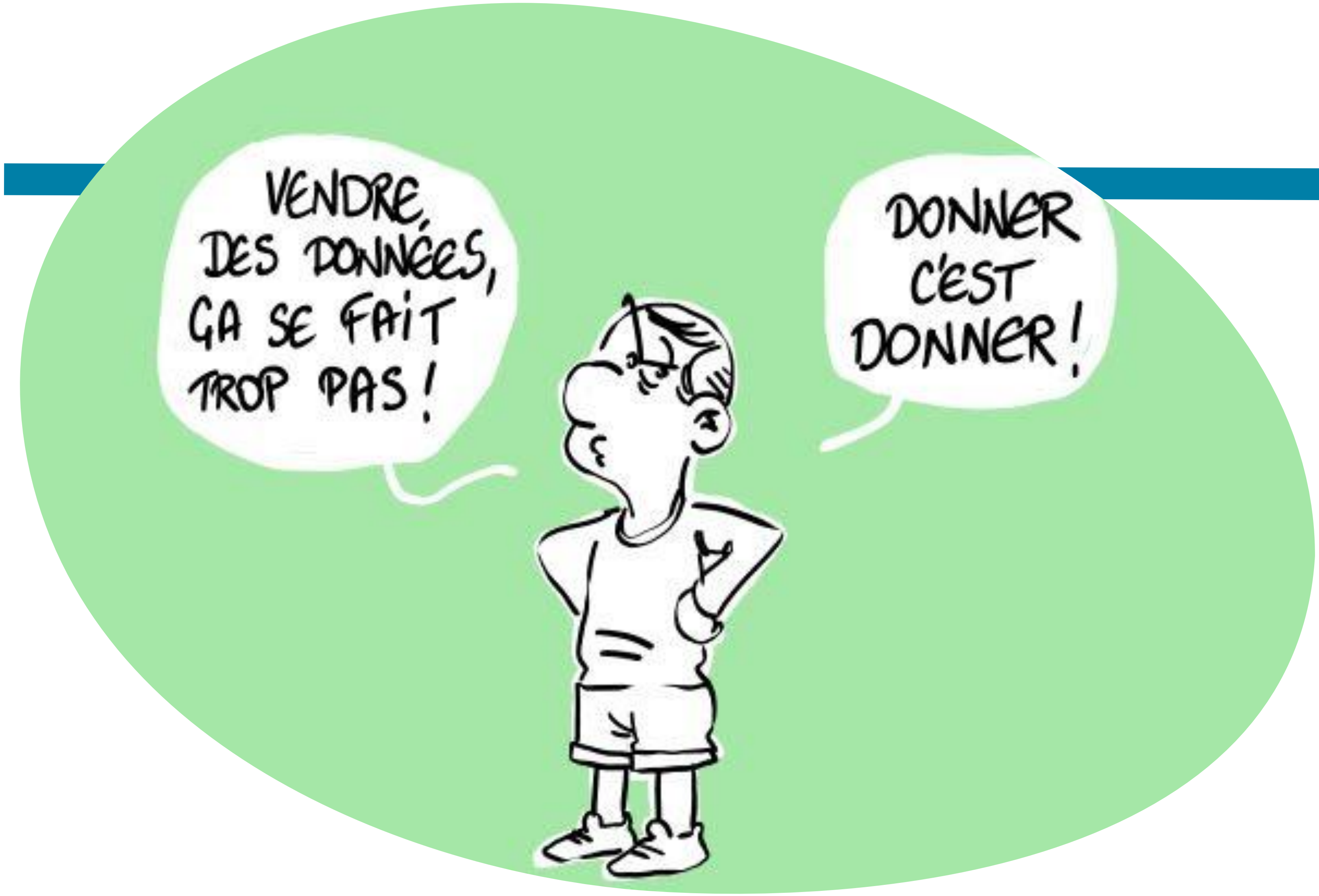
- OCT 2021 : Popularization MOOC on **AI by Example** (***AI360***)
  - In French (ANR THEME)
  - *Aquitaine region support with 4 start ups ?*

- OCT 2021 : French version of eBIHAR MS degree (with CAMPUS France) for Africa : ESATIC (Ivory Coast), ITU (Madagascar) along with their French Gradeos
  - Local « Digital connected campus » as third learning place

- OCT 2021 : M1 and M2 for BIHAR (2 years) at ESTIA (and on-line with eMiage with projects in Africa (Senegal) and Oceania (Tahiti)

- European PROPOSALS around ESTIA BIHAR:
  - DATA TOKI Erasmus proposal in 2020 with an annual event : The FORUM of the DATA ECONOMY in Biarritz
  - ERASMUS MUNDUS Master project in 2021 based upon blended learning (and moocs) with University of Siena (Italy) and Bilbao (Spain)
  - DEEP BRIDGE research project on Scanner image analysis for brainstroke detection (with Hospital of Nice, ECRIN, Inria, Eurecom, Siena Univ,..)
  - LIMAD project in Senegal (UVS) with AI blended curriculum from BAC to BAC+8 with Univ of Bordeaux

EXTRA SLIDES ☺



« DATA HUMANUM ES »!

JOB OPPORTUNITIES and bottom-up approach for Gradeos：130,000 engineers needed by 2025

https://insights.dice.com/2020/11/16/artificial-intelligence-a-i-job-trends-important-to-watch-in-2021/*

Occupation	Total Job Postings	Job Postings Requesting Skill(s)(#)	Job Postings Requesting Skill(s)(%)	Projected Growth Within Occupation (%)	Associated Education Level
Software Developer / Engineer	1,088,223	33,395	3.1%	30.7%	Bachelor's degree
Data Scientist	42,050	28,611	68.0%	19.0%	Bachelor's degree
Network Engineer / Architect	177,097	7,830	4.4%	6.5%	Bachelor's degree
Data Engineer	38,571	6,928	18.0%	11.5%	Bachelor's degree
Data / Data Mining Analyst	89,573	6,257	7.0%	9.3%	Bachelor's degree
Computer Systems Engineer / Architect	187,204	4,937	2.6%	9.3%	Bachelor's degree
Researcher / Research Associate	82,175	4,511	5.5%	27.5%	Bachelor's degree
Product Manager	101,358	3,940	3.9%	10.1%	Bachelor's degree
Database Architect	50,583	3,011	6.0%	9.3%	Bachelor's degree
Business / Management Analyst	306,954	2,448	0.8%	14.3%	Bachelor's degree

Source: Burning Glass (JOBS that require ML/DL)

eBIOAR

**The use of artificial intelligence (A.I.) and machine learning (ML), technologies that help people and organizations handle customer personalization and communication, data analytics and processing, and a host of other applications continues to grow. An IDC report found three-quarters of commercial enterprise applications could lean on A.I. by next year alone, while an Analytics Insight report projects more than 20 million available jobs in artificial intelligence by 2023.*

Due to A.I. and ML's transformational reach, specialists with the right skills could find themselves with job opportunities across a wide range of industries.

« LES » Sciences (Science des DATA)

- CONCEPTS
- METHODES
- OUTILS

SIM / MIS ? (Mobiquitous Information system)

- *« MIS supports eniantrodomic** holomophic infostructures for commonactors SURFING our ROR* future in smart places among smart objects with smartphones in a bottom-up serendipity (4th) paradigm of science based upon DATA »*

- « Système d'Information Mobiquitaires/Massives » (SIM) : *« Les SIM correspondent à des infostructures holomorphes eniantrodiomiques** permettant aux communacteurs de surfer en mode ROR* les écounomènes intelligents dans un 4ième paradigme oblatif de sérendipité de la science des DATA »*

* ROR : return on relationship (>> ROI)
**Eniantrodomia (Greek) : interdependy of opposites »

Deep Learning holomorphism

Holomorphic set : « *ALL is DEEP LEARNING* »
cf Vitaly Vanchurin (August 2020)

Cf MONADS (Leibnitz)

The world as a neural network

Vitaly Vanchurin

Department of Physics, University of Minnesota, Duluth, Minnesota, 55812
Duluth Institute for Advanced Study, Duluth, Minnesota, 55804

E-mail: vvanchur@d.umn.edu

Abstract.
We discuss a possibility that the entire universe on its most fundamental level is a neural network. We identify two different types of dynamical degrees of freedom: "trainable" variables (e.g. bias vector or weight matrix) and "hidden" variables (e.g. state vector of neurons). We first consider stochastic evolution of the trainable variables to argue that near equilibrium their dynamics is well approximated by Madelung equations (with free energy representing the phase) and further away from the equilibrium by Hamilton-Jacobi equations (with free energy representing the Hamilton's principal function). This shows that the trainable variables can indeed exhibit classical and quantum behaviors with the state vector of neurons representing the hidden variables. We then study stochastic evolution of the hidden variables by considering D non-interacting subsystems with average state vectors, $\bar{\mathbf{x}}^1$, ..., $\bar{\mathbf{x}}^D$ and an overall average state vector $\bar{\mathbf{x}}^0$. In the limit when the weight matrix is a permutation matrix, the dynamics of $\bar{\mathbf{x}}^\mu$ can be described in terms of relativistic strings in an emergent $D+1$ dimensional Minkowski space-time. If the subsystems are minimally interacting, with interactions described by a metric tensor, then the emergent space-time becomes curved. We argue that the entropy production in such a system is a local function of the metric tensor which should be determined by the symmetries of the Onsager tensor. It turns out that a very simple and highly symmetric Onsager tensor leads to the entropy production described by the Einstein-Hilbert term. This shows that the learning dynamics of a neural network can indeed exhibit approximate behaviors described by both quantum mechanics and general relativity. We also discuss a possibility that the two descriptions are holographic duals of each other.

« BIG FIVE » in the data *wild world* ?

GAFAM…

« BIG FIVE » in the data *wild world* ?

Google and (Androïd-based) SCREENS

INSTITUTE OF TECHNOLOGY

Android and Smart TV

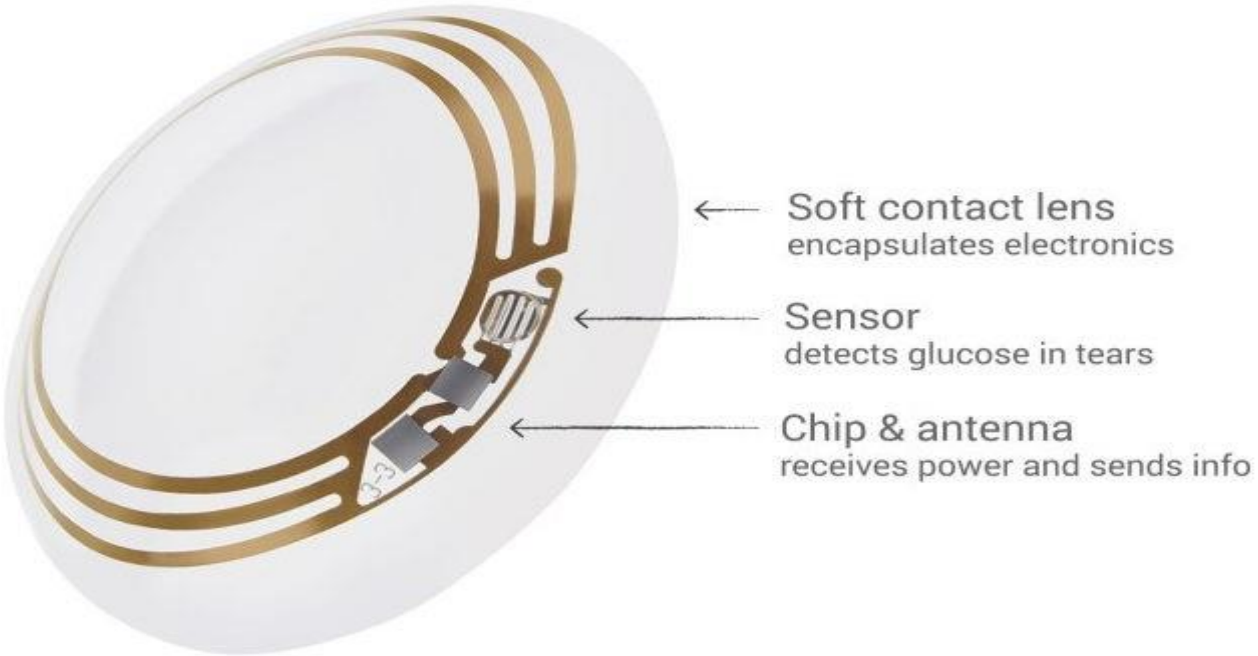
ANdroïd and Smart Home (NEST)

Android and Smart Car, Smart Glasses, Smart Watch…

e-BIMAR

Google and geo_loc advertising

Google and health



Google Strategy

1. « KNOWING YOU »

2. « multi-screen addiction »

   - (Google Home/assistant, Google car, Google Health..)

TOP DOWN approach for semi-structured DATA stores

- OPEN DATA

- WEB DATA (Semantic web)
  - RDF (Resource Description Framework) paradigm

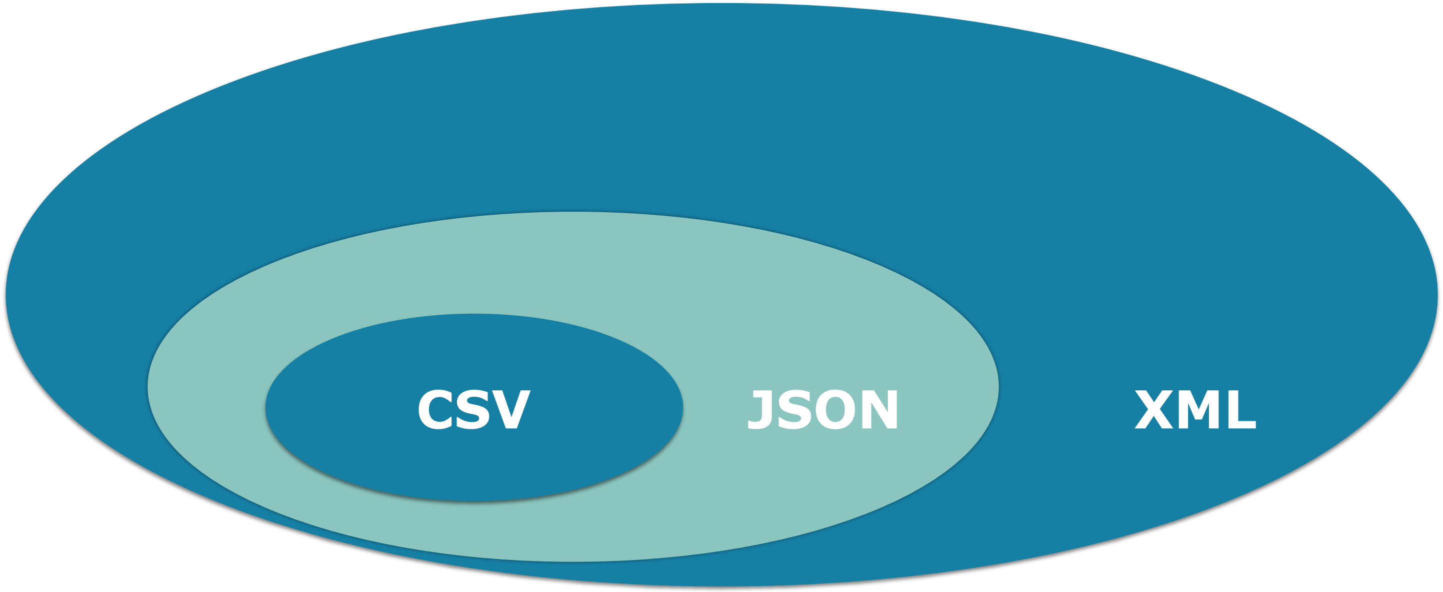
OPEN DATA formats

PDF for documents

For DATA :

- **CSV** (Excel)
- Web standards for publication and sharing
  - HTML (HTML5), **XML,** RDF
- Web standards for syndication
  - RSS, Atom, **JSON**

OPEN DATA : CSV, JSON, XML



CSV (Comma Separated Value) for flat files (1)
JSON (Java Script Object Notation) for hierarchical documents (2)
XML (eXtensible Markup Language) for (1), (2), namespaces,…

CSV, JSON and XML (Examples)

#CSV example
First, Name, Course title, date
« Serge », « Miranda », « From data bases to Big Data », « 2020 »

// JSON example
{« First »: « Serge », « Name »: « Miranda »,
« course »: {« title »: « From data bases to BIG DATA », « date » : « 2020 »}}

<!- XML example -->
<xml> <professor>serge Miranda</professor>
<list>
<course>From data bases to BIG DATA</course>
<date>2020</date>
</list>
</xml>

DATA WEB (semantic web)

*« I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web — (the content, links, and transactions between people and computers). ..A « **Semantic Web** », which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The « intelligent agents » people have touted for ages will finally materialize »*

TIM Berners Lee (2001, Weaving the web)

- WEB evolution :
  - **Network of PAGES ➔**
  - **Network of structured documents (XML) ➔**
  - DATA WEB/**Network of DATA** (RDF)➔
  - **Semantic web** (**Linked RDF**) **< W3C>**

RDF (*Resource Description Framework*)

- **Defined by W3C (January 15th, 2008)**
- Derived from XML
- **URI** for resource identification
  - Web page (identified by URL)
  - Web Service
  - XML document fragment
  - Any object (even physical) having collected DATA

DATA in RDF

RDF triples **to describe WEB resources**

(:serge: insureFLIGHT:AF100)

(:Peter:insureFLIGHT:AF110)

(:AIRBUSA320:isusedinFLIGHT: AF100)

(:Paul:ispassengerinFLIGHT:AF100) …

Note :

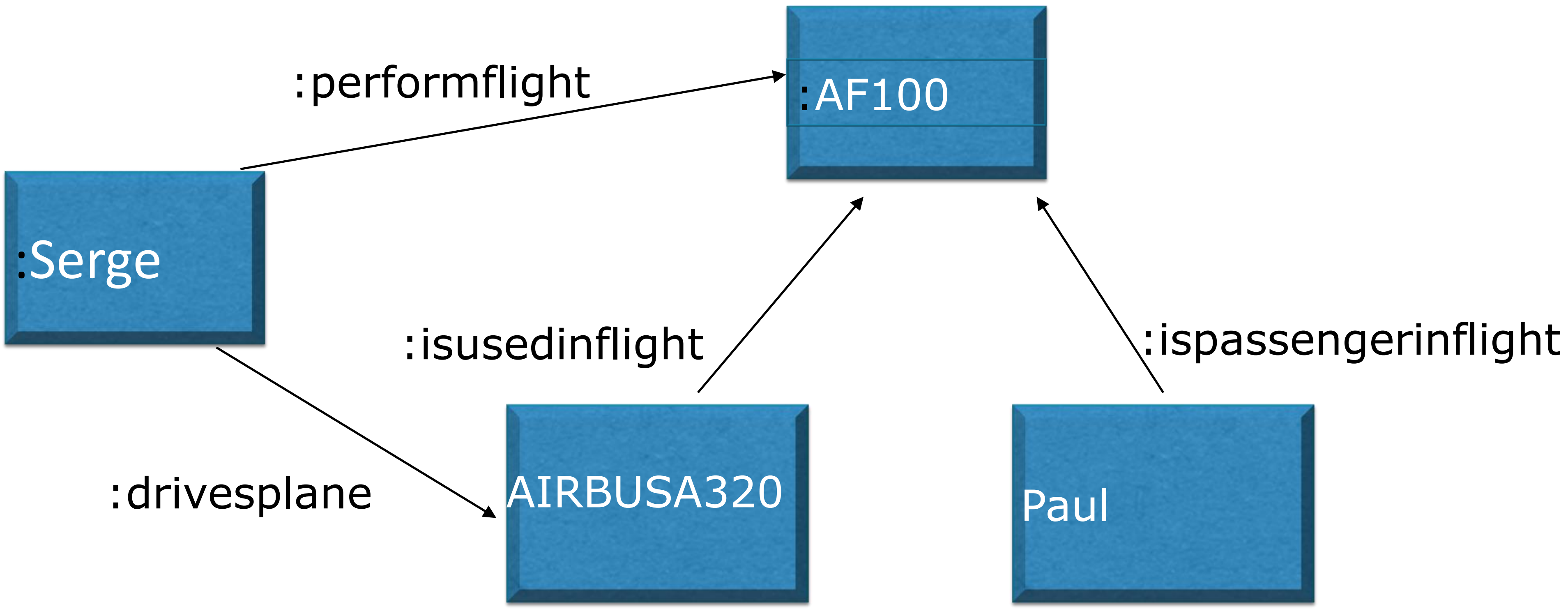
A RDF triple<S.P.O>

- is a *fact* in 1st order predicate logic

- P(S,O) with *P Predicate, S Subject and O object*

Example : INSUREFLIGHT (Serge, AF100)

RDF graph (Example)



R language : esperanto in DATA SCIENCE
(top 10 language in 2015 by IEEE; top 5 in 2016….)

- **R** is OPEN SOURCE (GNU GPL) for STATISTICA ANALYSIS on Linux, Windows, MacOS,.. with 2 major assets :
  - Social network
  - CARTOGRAPHY

- Created in 1993 by **R**oss Ihaka and **R**obert Gentleman from BELL (derived from S and SCHEME)
  - Writtent in C++, Java , Fortran and C

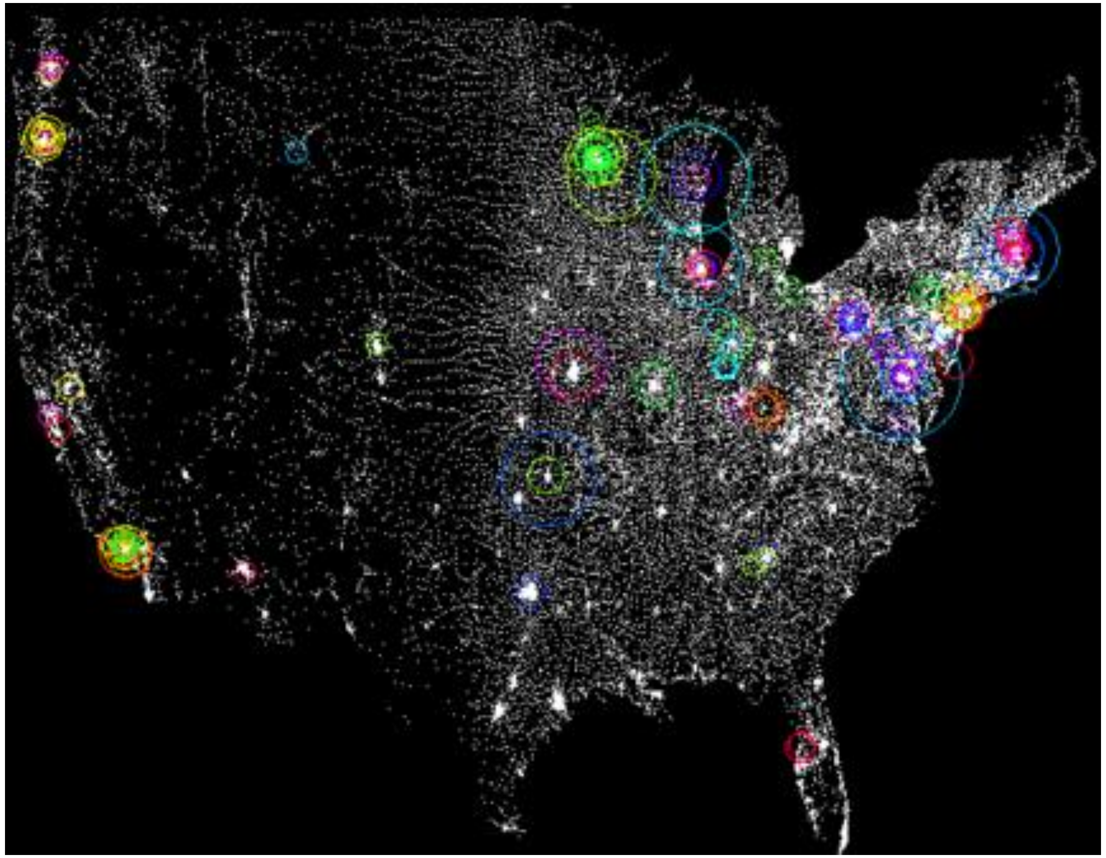
- Thousand of open libraries « **PACKAGES** » beyond basic statistics : from social network analysis to Deep Learning
  - **DATA FRAMES (matrices)**

- R Interface with every major DBMS ➔ Enterprise adoption Oracle, Microsoft, IBM, Teradata, Postgres, MySQL,.. With RMySQL, ROracle, RPostgreSQL ,…

R cartography example*

Credit card fraud scheme featuring time, location, and loss per event, using R :

Each circle is a fraudulent transaction in one particular fraud case, over several months. Circle radius represents dollar amount. Color represents recency, from blue (old) to red (new). The fraud spread from the East to the West coast, as you can tell by the colors.



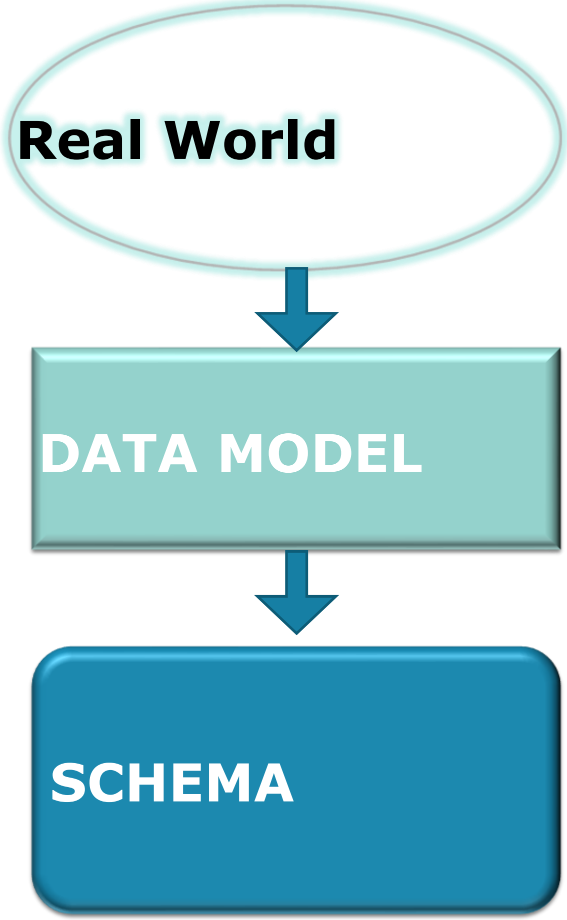
*GRANVILLE 2008

http://www.analyticbridge.com/photo/2004291:Photo:1417?context=featured

Top Down approach with SQL/ODMG

TOP DOWN approach for DATA **STRUCTURATION**

→ pre-definition of a fixed schema

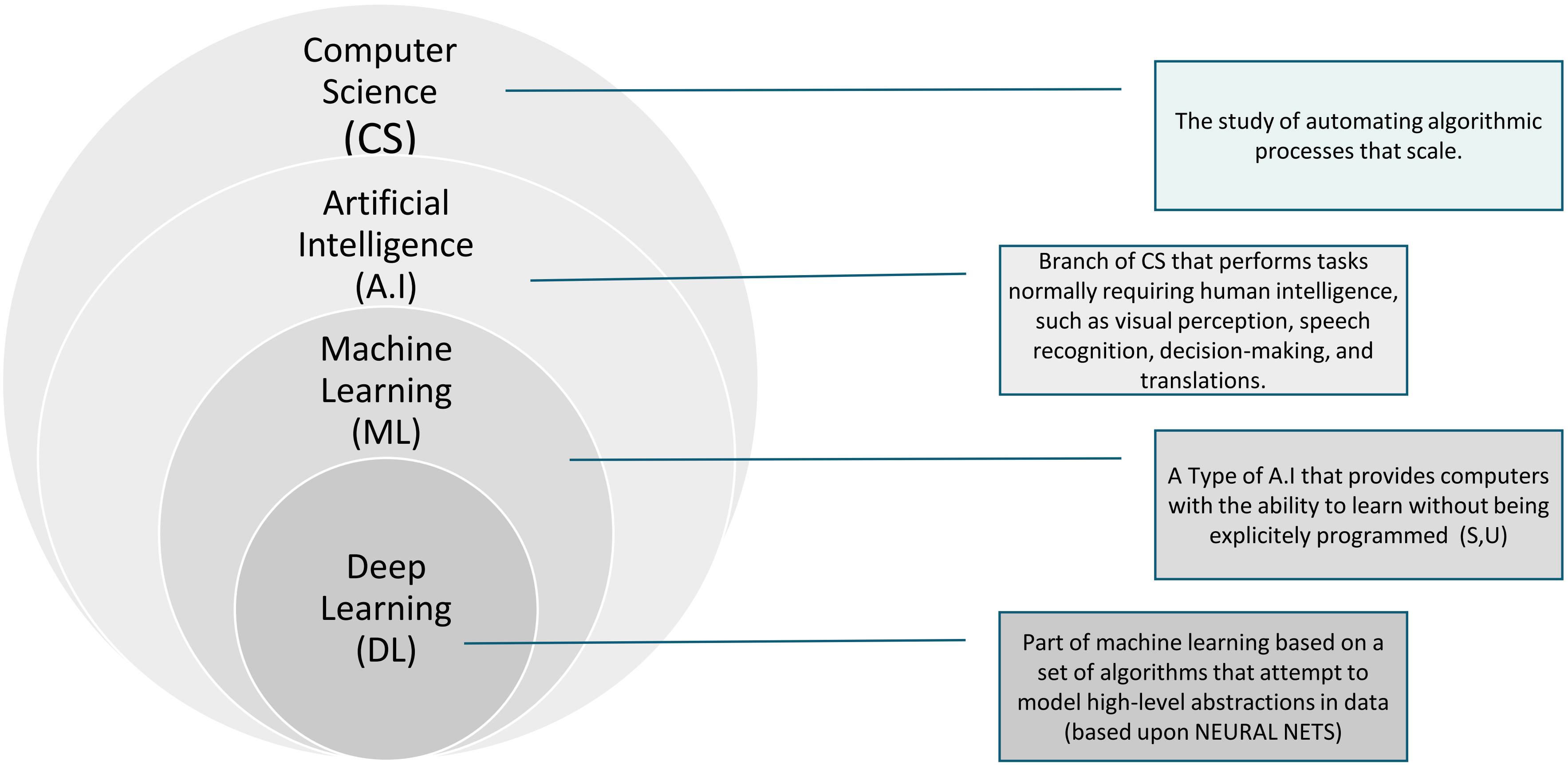


Gartner 's quadrants on DBMS (2019)

Gartner magic quadrants on DBMS in the cloud (Nov 2020)

Gartner has positioned Google as a Magic Quadrant Leader with AWS, Microsoft and Oracle

The case of Deep Learning (« neural nets »)



AI birth in Darmouth (August 1956)

1943 : **CONNECTIONISM** with Warren McCulloch, Walter Pitts, neurologists : **1st Neurone modelling**

1950 (oct) **COGNITIVISM** with Alan Turing : « ***Computing machinery and intelligence*** » : *Can machines think?*

1951 : **Snarc** (Stochastic Neural Analog Reinforcement Calculator) by Marvin Minsky (MIT) based upon Donald Hebb principles written in 1949 with learning capability from digital neurons

August 1956 : AI birth in Darmouth College with Marvin Minsky from MIT (1st digital neural net)s) and John McCarthy (LISP creator and words « ARTIFICIAL INTELLIGENCE »*)

* AI by John McCarthy : « *computing programs doing human tasks* »

AI history < cont’>

ESTIA
INSTITUTE OF TECHNOLOGY

- 1957 : Neural nets with Franck ROSENBLATT (Psy, Univ Cornell) who built PERCEPTRON (1 layer of neurons) inspired by Donald Hebb theory
- Multi-layer perceptron (MLP) for group classification which are non-linear
- Artificial Neuron = mathematics **functio**n

1959 : **reinforcement machine learning** by Arthur Samuel

Then 2 long AI winters…until 2012 (Imagenet)

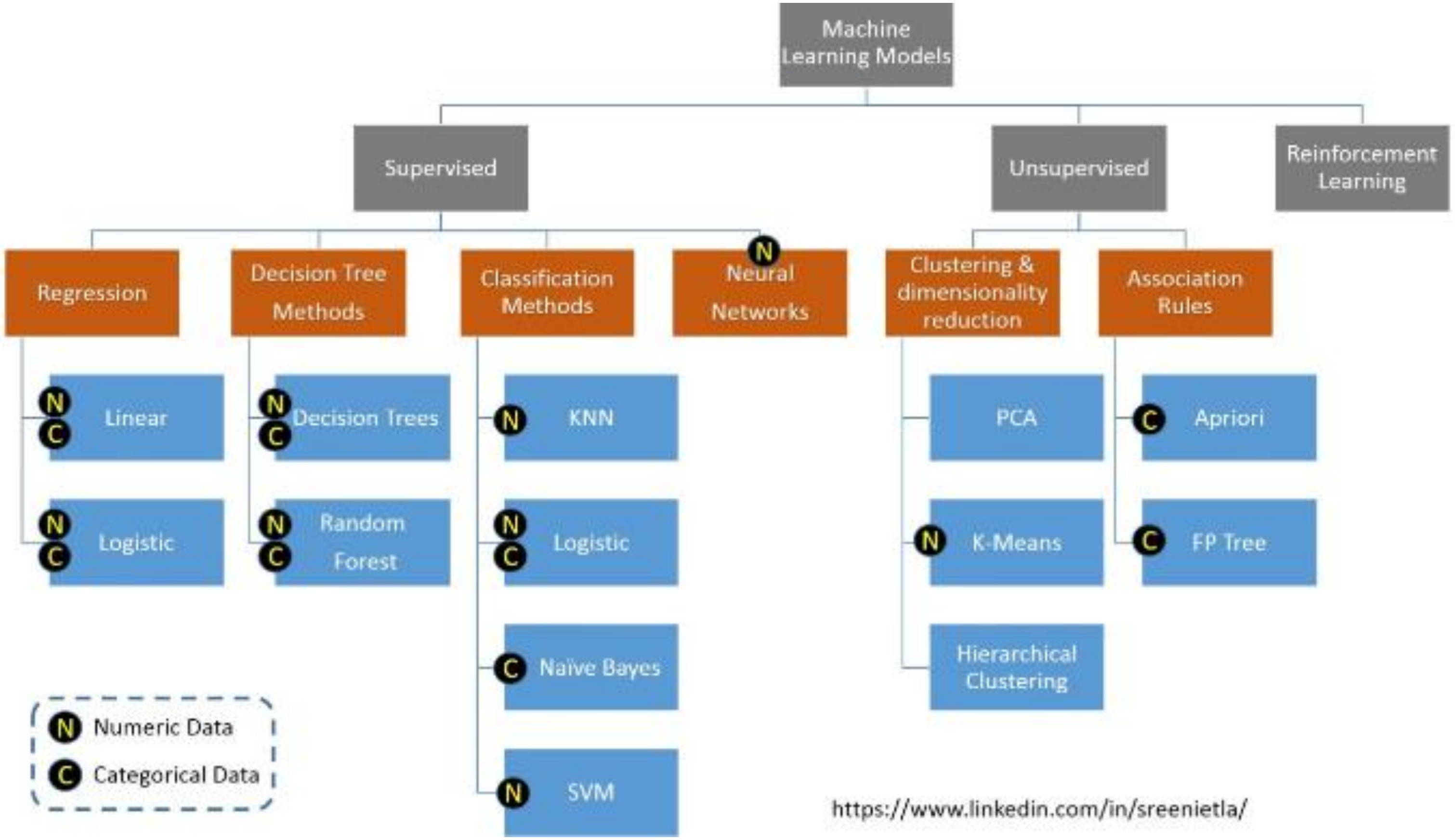
2019 : Turing Award for 3 researchers on digital neural nets & Deep Learning (Y.Bengio, G. Hinton and Y. Le Cun)

*

eBIMAR

Decoding Machine Learning

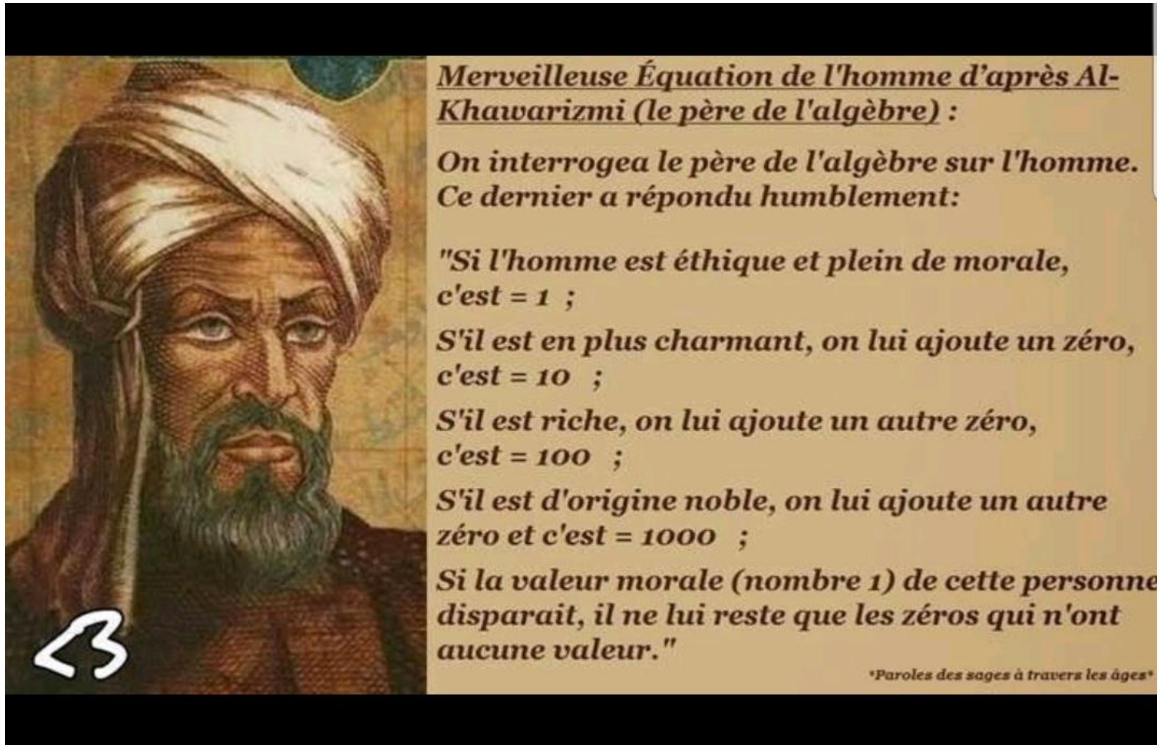
https://www.linkedin.com/pulse/decoding-machine-learning-sreenivasa-etla/



ESTIA
INSTITUTE OF TECHNOLOGY

Digital Algorithms ?

- A recipe/method ☺
- **AL KHOVARISMI (Persia, 9th century) & algebra !**
- « human algorithm »
  - LIVING is algorithmic!
- **« digital algorithms »**
  - Recommandation algorithms (ITTT paradigm)
  - Evolutionary algorithms (ML and Deep Learning)
    - Autonomous



eBIHAR

Treshold and *activation/step function*

- ACTIVATION (STEP) function
  - Applied to the weighted sum of INPUTS to get an OUPUT
  - New parameter to be added to the sum before the activation function : BIAS of the neurons
  - Learning methods for WEIGHTS (*supervised, non-supervised, reinforced*):optimization methods

- Types of **activation function** (for neurons differences)
  - LINEAR function
  - Treshold function
  - Radial function
  - Stochastic function

Core components of a DNN

- **Parameters :** We learned that parameters relate to the *x* parameter vector in the equation ***Ax = b*** in basic machine learning. Parameters in neural networks relate directly to the weights on the connections in the network.
- **Layers**
- **Activation functions**
- **Loops (output→ input) or NOT**
- **Loss functions**
- **Optimization methods** (such as *gradient descent* to find good values for the weights)
- **Hyperparameters** (layer size, number of neurons per layer,..)

Cost function and back-propagation of the gradient in CNN

- Function selection?
  - 2 parameters : polynomial of degree 1 (*straight line)*
    - Linear function y = f(x,a) = ax + b (a : slope)
  - 3 parameters : polynomial of degree 2 (*parable*)
- **GRADIENT** : Direction of the greatest slope = **gradient of the cost function**
  - First derivative of f(x,a) = a
    - Partial derivative : N-variable function derivative for one variable only
    - The vector built with partial derviatives is the gradient
- **Cost function** : *square of the difference between the generated output and the expected output Y*
  - If one output : Cost = (Y-f(x,a))**2
  - If n outputs : VECTOR with n cost values
- Learning mechanism : minimization of the cost function by *gradient descent*

Major DNN architectures

INSTITUTE OF TECHNOLOGY

- Unsupervised Pretrained Networks (UPN) **:**
  - *Autoencoders, Deep Belief Networks (DBNs), Generative Adversarial Networks (GANs)*
- Recurrent Neural Networks (texts,..)
- Recursive Neural Networks
- Multi-layered perceptron (MLP)

- **Convolutional* Neural Networks (CNNs)**
  - **Well suited to object recognition (image, sound, etc.)**

* ***Convolution*** is a mathematical function performed on two functions (used in digital signal processing); images could be treated as 2-dimensional functions

Machine learning and *Convolutional** Neural Nets

- ***AX = B*** in basic machine learning with the matrix A , the parameter vector x , to get output column vector b with **X VECTOR for weights** : {w1j, w2j, ..wnj}

- Machine learning, *statistical learning* (*neuron learning : weights*?) with data-centrics approach :
  - Supervised (labelled data) :
    - correct human guidance and B exists to find F (the prediction function)
  - Unsupervised :
    - no B➔ classification (*clustering*)
  - Hybrid
  - Reinforced : millions of probes to learn
  - **Adversarial*** : uncertainty prediction

- * both invented by Yann LeCun , Facebook Artificial Intelligence Research (FAIR) cf Harvard Business Review, Dec 2018 ***< TURING AWARD in 2019>***

SEMI-STRUCTURED DATA in RDF

TRIPLEs (OBJECT-PREDICATE-VALUE)

- To describe WEB resources

  (:serge: performflight:AF100)

  (:Peter:performflight:AF110)

  (:AIRBUSA320:isusedinflight: AF100)

  (:Paul:ispasengeroflight:AF100) …

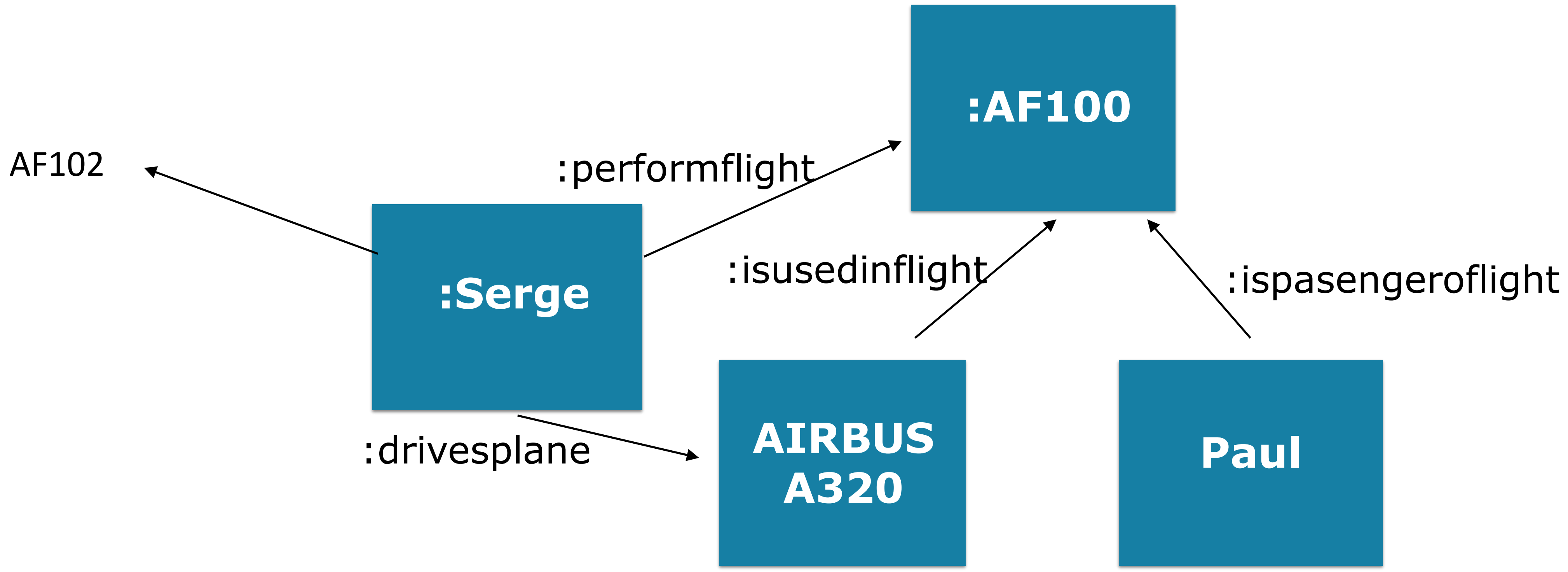
- Note :

  One triple RDF <S.P.O> is a fact in 1st-order predicate logic : P(S,O) with

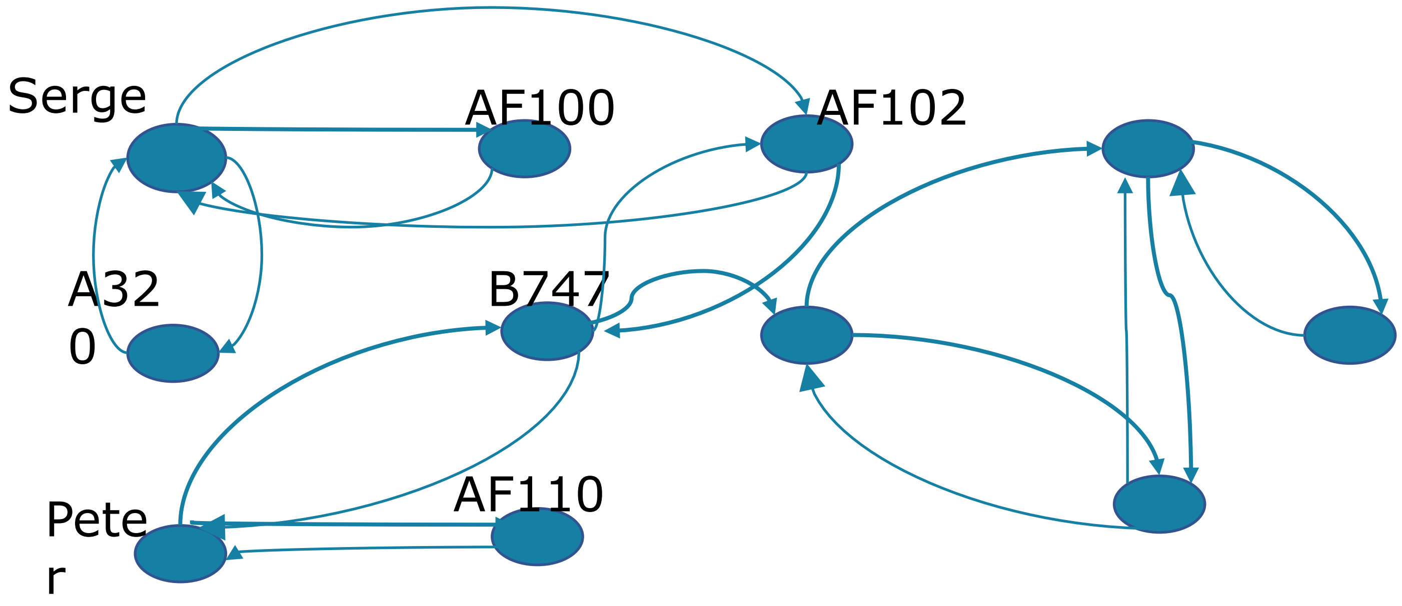
  P: Predicate, S Subject et O object

  Example : Performflight (Serge, AF100)

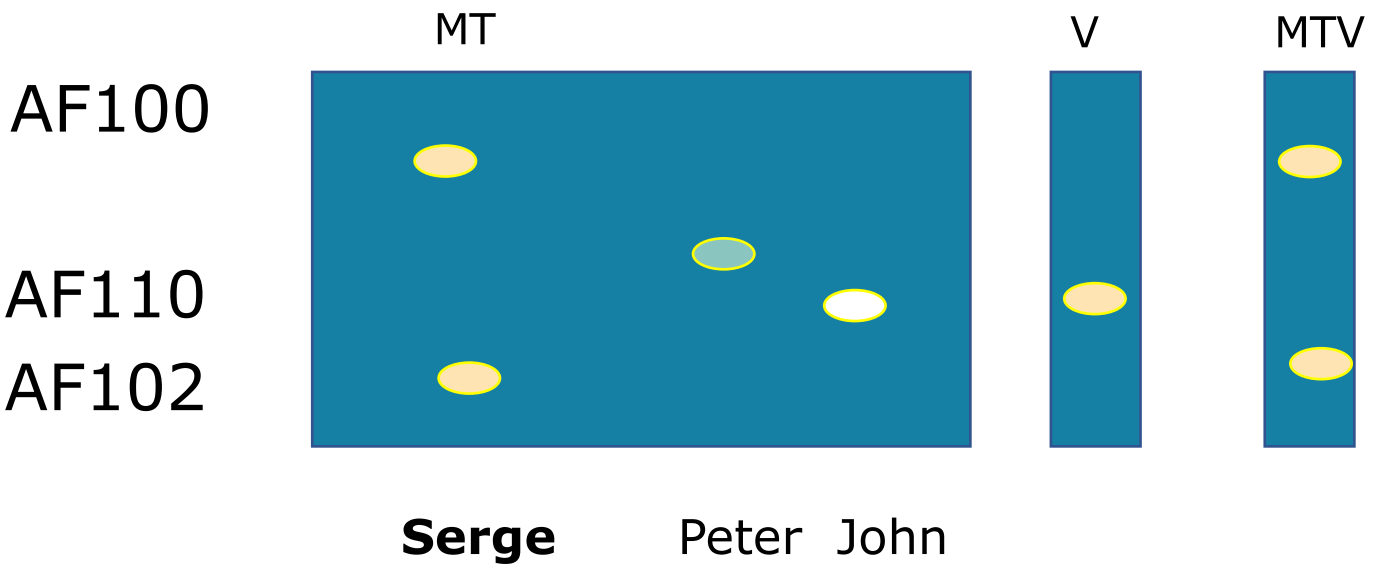
RDFS graph (Example)



NoSQL **graph** for unstructured data



NewSQL matrix



Sparse matrices and linear algebra!

OLAP

- Fact table (STAR or SNOWFLAKE) / Dimensions
- Generating **Cube operator** (with **ALL** value)
- Aggregation operations
  - Slicing
  - Dicing
  - Roll-up
  - Drill-down/up

SQL and No SQL/new SQL

- SQL focus on SET THEORY for **TRANSACTIONS**

- NoSQL and NewSQL focus on Graph theory and Matrix mathematics (linear algebra) for high-performance **DATA ANALYSIS and decision support** with mathematical properties such as : Associativity, commutativity, distributivity, identity, annihilator and inverses

Graph algebra*

- Some basic binary and unary operations
  - 3 binary : UNION, INTERSECTION, SYMMETRIC DIFFERENCE
  - 4 Unary :
    - Vertex Removal
    - Edge Removal
    - Vertex Identifying
    - Edge Contraction
- Search or Graph Traversal algorithm :
  - Such as breadth-first search (BFS), depth-first search (DFS), …

- *YII Haxhimusa. The Structurally Optimal Dual Pyramid and its Application in Image Partitioning. Vienna University of technology, Faculty of Informatics, Institute of Computer Aided Automation, Pattern Recognition and Image Processing Group. PhD Thesis, Chapter 2, 2006*

- *+ GQL (Graph Query Language) proposed to standardization by Neo4J, Oracle… with MATCH operator in 2018*

Matrices – Linear Algebra (LA)

- Matrices are common representation for computation
- common data representation for Machine Learning (ML) and Deep Learning (DL)
  - ML algorithms (such as Linear regression, Logistic regression and K-Means (Clustering) can be expressed succinctly using LA operators such as matrix multiplication and inversion*

*L. Chen, A. Kumar, J. Naughton, et J. M. Patel, « Towards linear algebra over normalized data », Proceedings of the VLDB Endowment, vol. 10, no 11, p. 1214 - 1225,August 2017

Linear Algebra for ML / DL

Operation	Arithmetic
Matrix Addition	C = A + B
Matrix Subtraction	C = A - B
Matrix Multiplication (Hadamard Product)	C = A o B
Matrix Division	C = A / B
Matrix-Matrix Multiplication (Dot Product)	C = A * B C(m,k) = A(m,**n**) * B(**n**,k)
Matrix-Vector Multiplication	c = A . v
Matrix-Scalar Multiplication	C = A . b
Transpose	C = A^T
Inversion	B = A^-1
Trace	tr(A)
Determinant	det(A)
Matrix Rank	rank(A)

sources :

http://wiki.fast.ai/index.php/Linear_Algebra_for_Deep_Learning

https://towardsdatascience.com/linear-algebra-cheat-sheet-for-deep-learning-cd67aba4526c

https://machinelearningmastery.com/matrix-operations-for-machine-learning/

Unifying underlying theory for BIG DATA

	Paradigm	Data structures	Math theory	Data model
SQL	VALUE	Relation/ Table	**SET theory**	Relational model (Codd's)
Semi-structured	PREDICATE/ VALUE	Class and RDF	**Graph theory**	RDF data model
NoSQL	Key-value & graph	Class Document (JSON)	**Graph theory**	Key/BLOB Key/document (JSON) Key/(column/ value)
NewSQL	Value & graph	Table	**Sparse matrices / Linear algebra**	Extended Relational model
R, Python interface	VALUE	Arrays	**Linear algebra**	Matrix

Some books of reference on Data bases

In English :

- **Chris Date « *An Introduction to data base systems* » (8th Edition), Addison Wesley *<the reference book on data bases>***
- E.F Codd (1990). « *The Relational Model for Database Management* » (Version 2). Addison Wesley Publishing Company. ISBN 0-201-14192-2. *<Codd's book>*
- M.Stonebraker et al « *Readings in data base systems* » *<The « red book »>* 5th Edition 1998, Morgan Koffmann
- S.Abitboul et al « *Foundation of data bases* » Addison Wesley *<data base theory>*

In French :

- **JL Hainaut « *Bases de données (Concepts, applications et développement)* », DUNOD, 4ième Edition, 2018**
- G. Gardarin « *Bases de Données* » Eyrolles, Version gratuite sur georges.gardarin.free.fr
- S. Miranda « *L'Art des Bases de données* » (3 Tomes), EYROLLES
- S. Miranda « *Bases de données : Architectures, modèles relationnels et objets, SQL3 et ODMG* », DUNOD, 2002

Some books of reference on Big Data

In English

- Rajendra Akerkar (Ed) “Big Data Computing” CRC Press, 2014
- Jules Berman “Principles of Big Data” Morgan Kaufman, 2013
- Joe Celko “”A Complete guide to NO SQL » Elsevier 2014
- W.CHU Editor « Data mining and knowledge Discovery for big data » Springer 2014
- Dan Mc Creary, Ann Kelly « Making sense of NO SQL » Manning 2014
- F.Provost, T Fawcell « DATA SCIENCE for Business » O’Reilly 2013
- Jordan Tigani, Siddartha Naidi « Google Bigquery Analytics » WILEY, 2014 (510 pages)
- *Mike Stonebraker, “New SQL: An Alternative to NoSQL and Old SQL for New OLTP Apps » ACM, June 2011*

In French

- R.Bruchez « Les bases de données NO SQL et le Big Data », Eyrolles 2015
- I.lemberger et al « « Big data et machine learning”, Dunod 2016
- C.Azencott “Introduction au Machine learning » Dunod 2018
- G.Grolemund « R pour les data science », Eyrolles 2017