Data Driven Business Introduction

Trainer’s name: Lali Soler

Email address:

lali.soler@eurecat.org

Contents

- Definitions and a bit of motivation:
  - Business Context
  - Big Data
  - Data Science
  - Machine Learning
- The “Data Science toolbox”
- The “Big Data” toolbox
- Market and industry
- Artificial Intelligence in the industry

eurecat
Centre Tecnològic de Catalunya

01

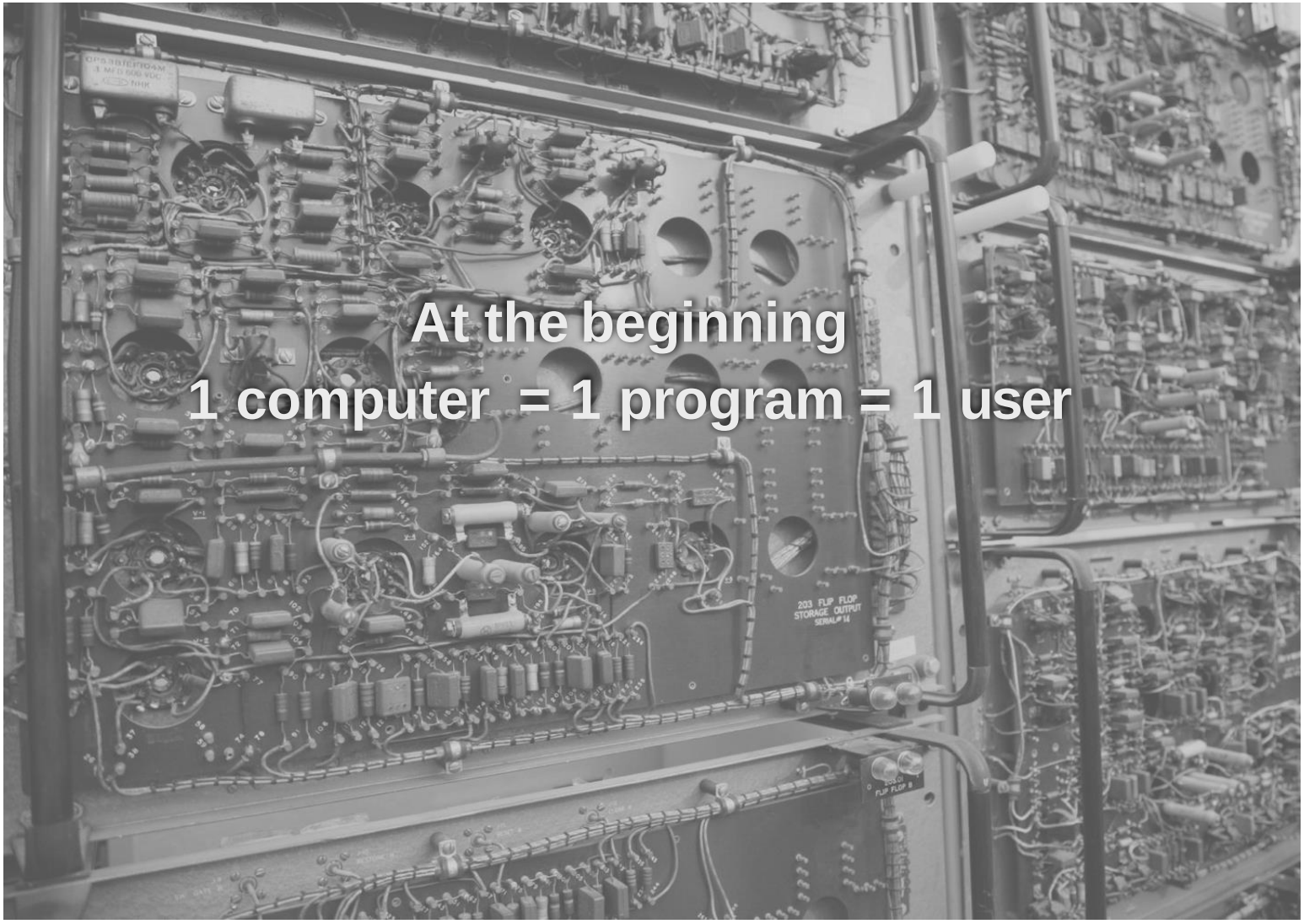
Introduction

3

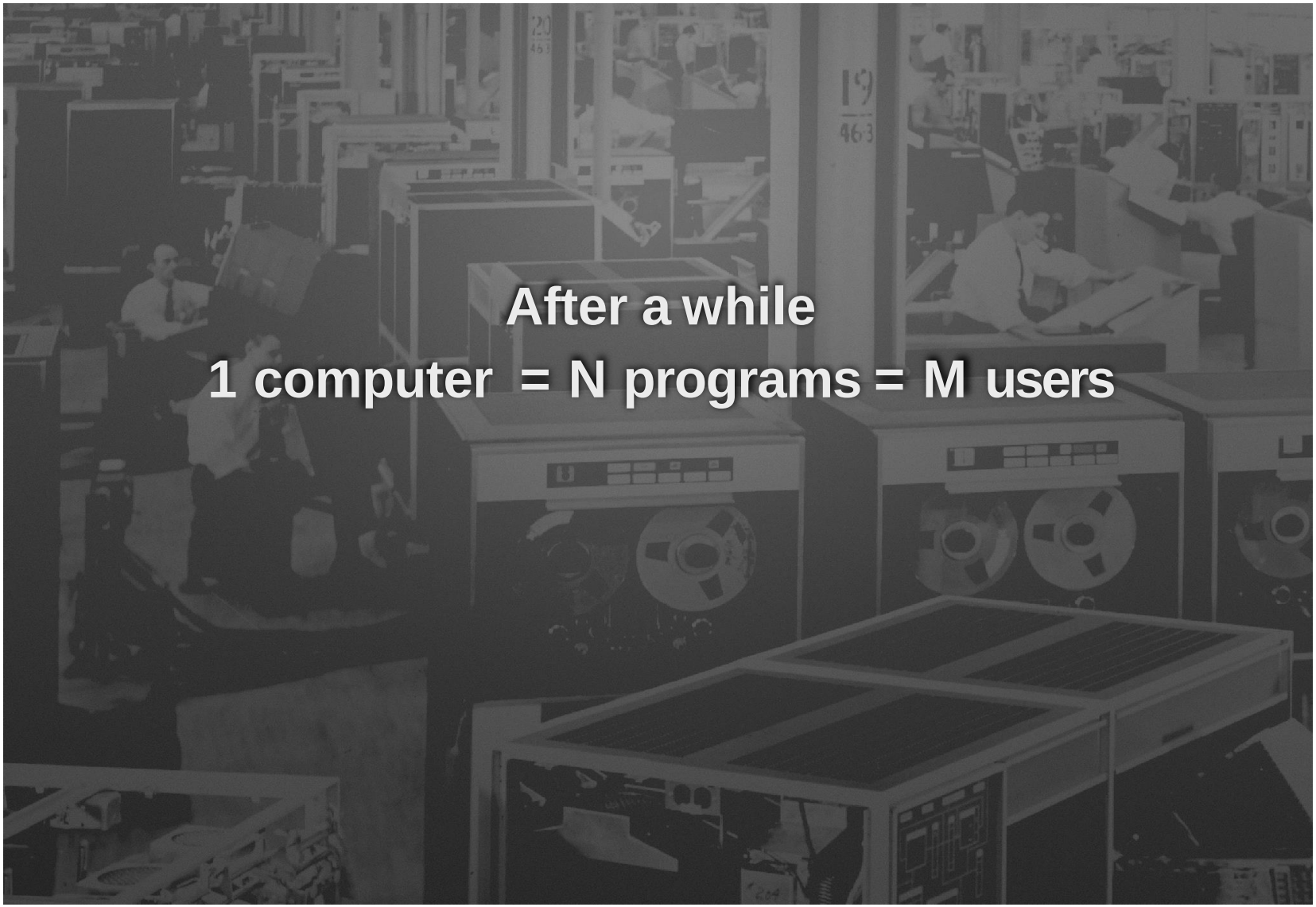
The value of Data

Making decisions based on data is nothing new. Now it is much easier, simply.

Why now?



Why now?



Why now?

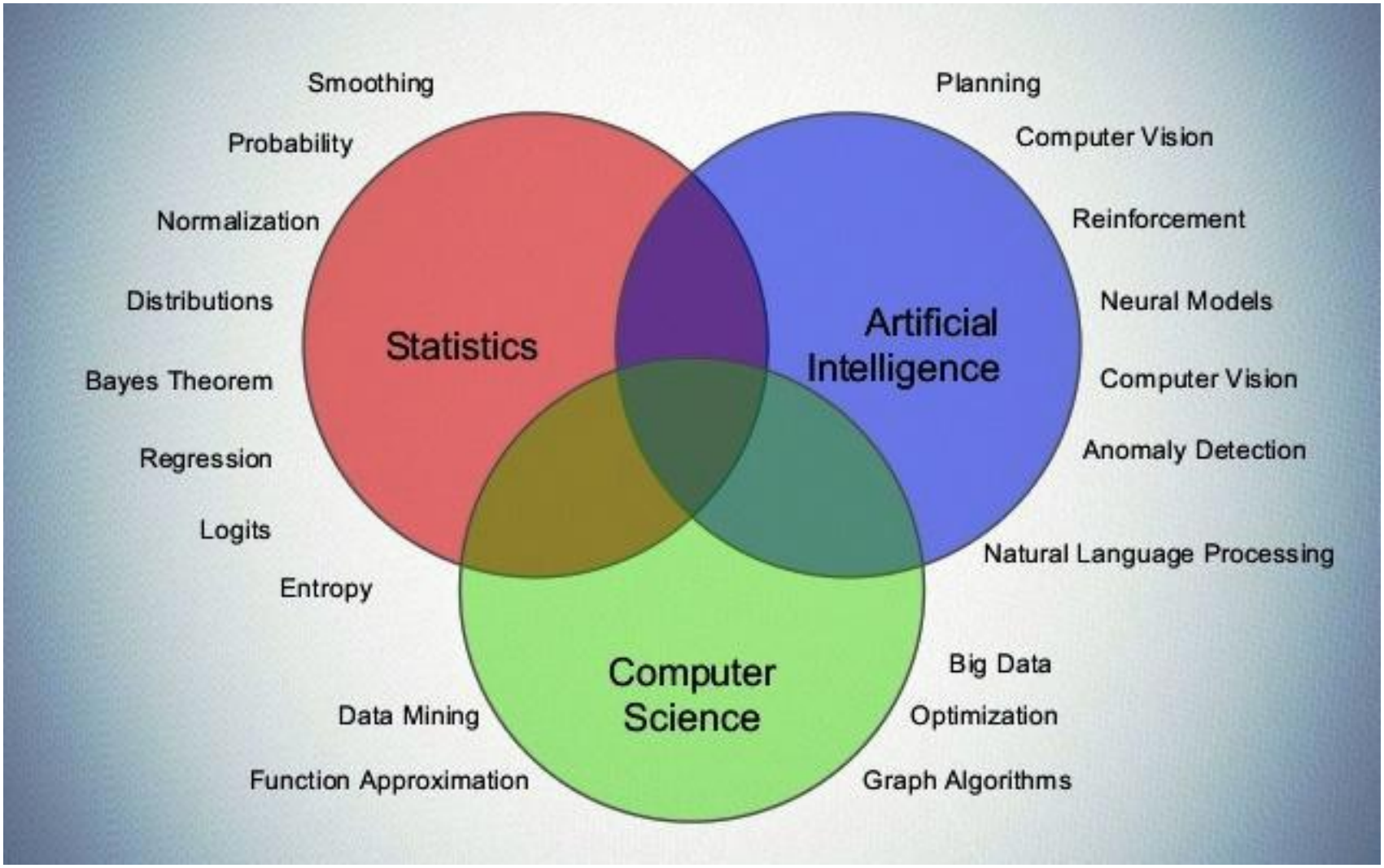
Why now?

A few years ago we reach the present situation.

From a user perspective:

M computers = N programs = 1 user

A bit of terminology



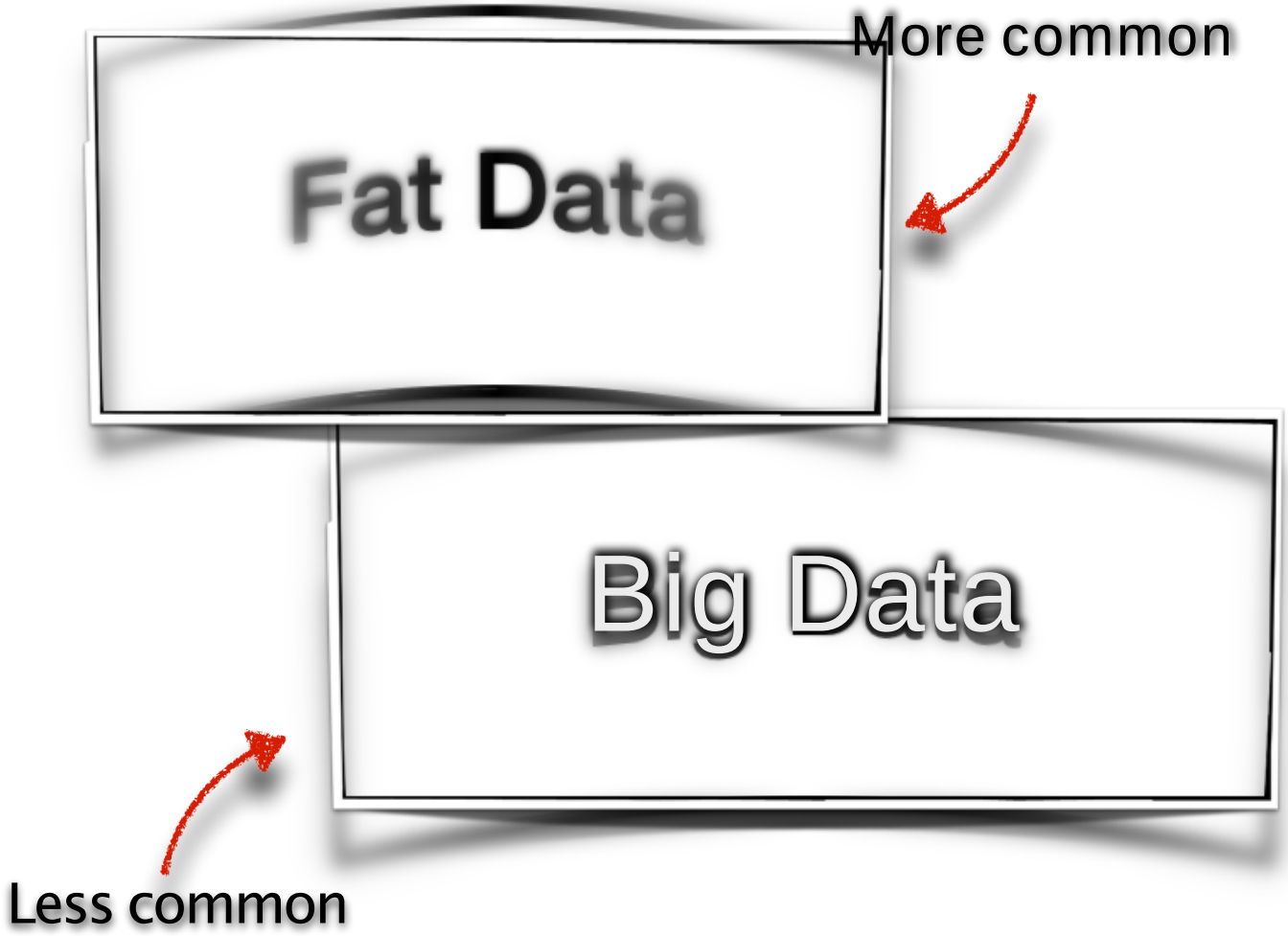
eurecat
Centre Tecnològic de Catalunya


Data Science
Big Data

eurecat
Centre Tecnològic de Catalunya

What is Big Data?

- For some people, they have big data when its size > 65536 x 256.
- In general we have big data when its size does not allow its storage and analysis in a big computer.


More common
Fat Data
Big Data
Less common

Big Data

eurecat
Centre Tecnològic de Catalunya

Wal-Mart handles over one million customer transaction per hour, the information is stored on a database sized in excess of 2.5 Petabytes ($2{,}0 \times 10^{16}$ bits).

By 2016 it is likely that a typical hospital will create 665 terabytes (5.32×10^{15} bits) of data a year.

eurecat
Centre Tecnològic de Catalunya

With a personal computer:

You can find an element in a 1 MB file in less than a second.
You can find an element in a 1 GB file in less than a minute.
You can find an element in a 1 TB file in less than sixteen hours.
You can find an element in a 1 PB file in less than two years.
You can find an element in a 1 EB file in less than two thousand years.

Big Data

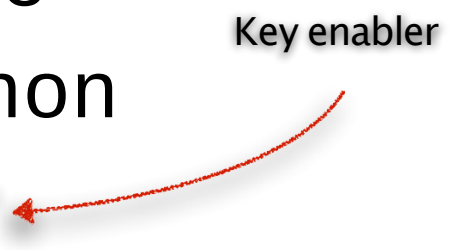
eurecat
Centre Tecnològic de Catalunya

Big data is more than size.

It is commonly characterized with four V

Big Data

The cloud is key to deal with the three V, but the main phenomenon behind Big Data is **datification**.



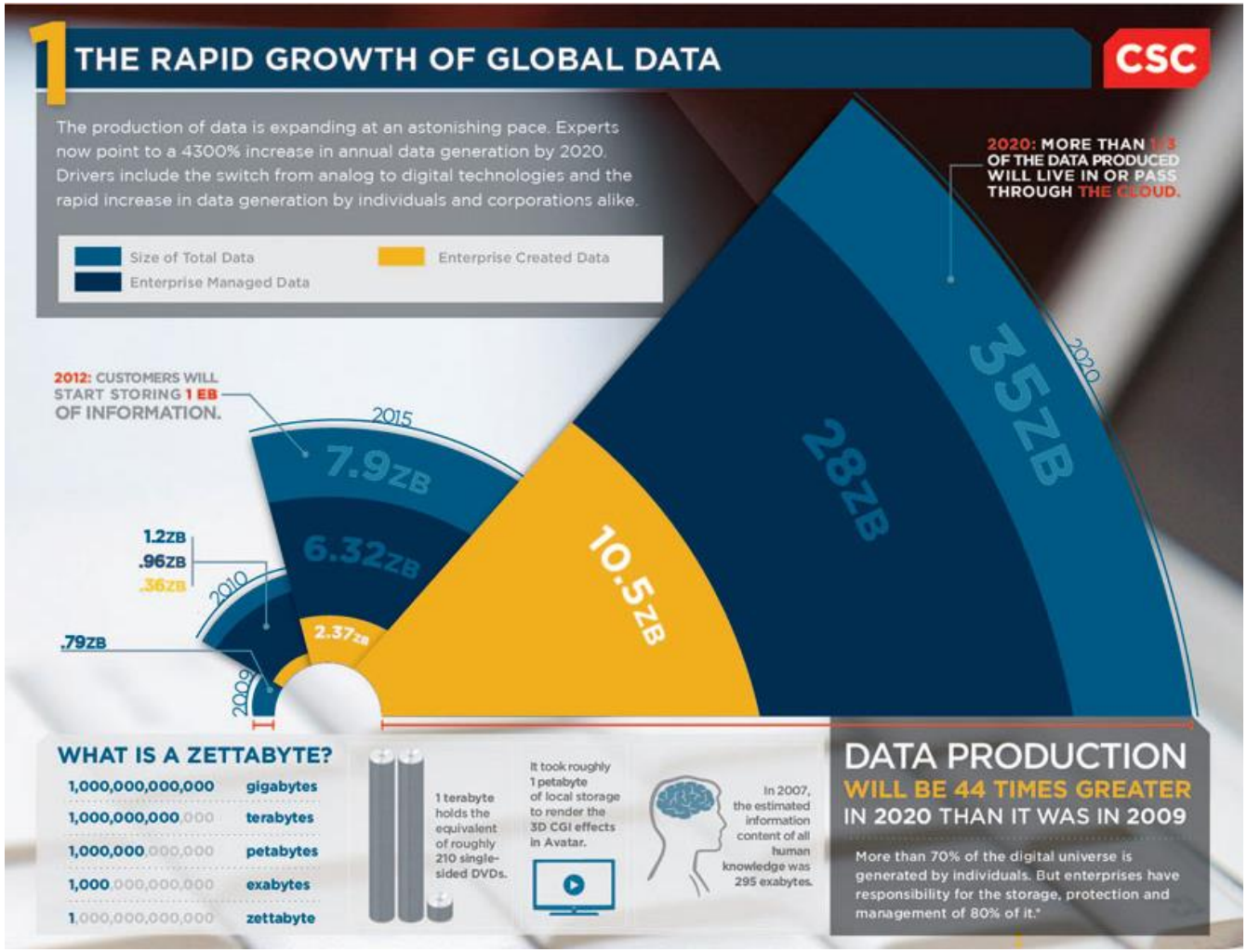
The three V are a consequence of it.

Big Data

eurecat
Centre Tecnològic de Catalunya

We are rendering into data many aspects of the world that have never been quantified before:

business networks
books I'm reading
location
physical activity
consumed food
purchases
physiological signals
straight thoughts
friendship
gaze
driving behavior


1
THE RAPID GROWTH OF GLOBAL DATA
CSC
The production of data is expanding at an astonishing pace. Experts now point to a 4300% increase in annual data generation by 2020. Drivers include the switch from analog to digital technologies and the rapid increase in data generation by individuals and corporations alike.
Size of Total Data
Enterprise Managed Data
Enterprise Created Data
2020: MORE THAN 1/3 OF THE DATA PRODUCED WILL LIVE IN OR PASS THROUGH THE CLOUD.
2012: CUSTOMERS WILL START STORING 1 EB OF INFORMATION.
2015
2020
2010
2009
1.2ZB
.96ZB
.36ZB
.79ZB
7.9ZB
6.32ZB
2.37ZB
10.5ZB
28ZB
35ZB
WHAT IS A ZETTABYTE?
1,000,000,000,000 gigabytes
1,000,000,000,000 terabytes
1,000,000,000,000 petabytes
1,000,000,000,000 exabytes
1,000,000,000,000,000 zettabyte
1 terabyte holds the equivalent of roughly 210 single-sided DVDs.
It took roughly 1 petabyte of local storage to render the 3D CGI effects in Avatar.
In 2007, the estimated information content of all human knowledge was 295 exabytes.
DATA PRODUCTION
WILL BE 44 TIMES GREATER
IN 2020 THAN IT WAS IN 2009
More than 70% of the digital universe is generated by individuals. But enterprises have responsibility for the storage, protection and management of 80% of it.*

Big Data

eurecat
Centre Tecnològic de Catalunya

Information comes from:

- Corporate Data Bases (structured information). Unstructured information in documents, Wikipedia, textbooks, journals, blogs, tweets, etc.
- Images in the web, public cameras, phones, TV, YouTube, etc.
- Public APIs: smart cities, government, search engines, etc.
- Sensor Data: GPS, accelerometer, physico- chemical sensors, sociometric sensors, super- colliders, telescopes, etc.

Big Data

eurecat
Centre Tecnològic de Catalunya

There are several Big Data flavors:

- Big multidimensional arrays (homogeneous data).
- Big tables (structured data).
- Big text.
- Big image.
- Big sound.
- Big sequential data (sensors, tweets, etc.)

Big Data

eurecat
Centre Tecnològic de Catalunya

There are several problems:

- ETL (Extract, Transform, Load)
- BI/Analytics (Think you can do in SQL)
- **Advanced Analytics.**
- **Machine Learning.**
- Visualization.

Data Science

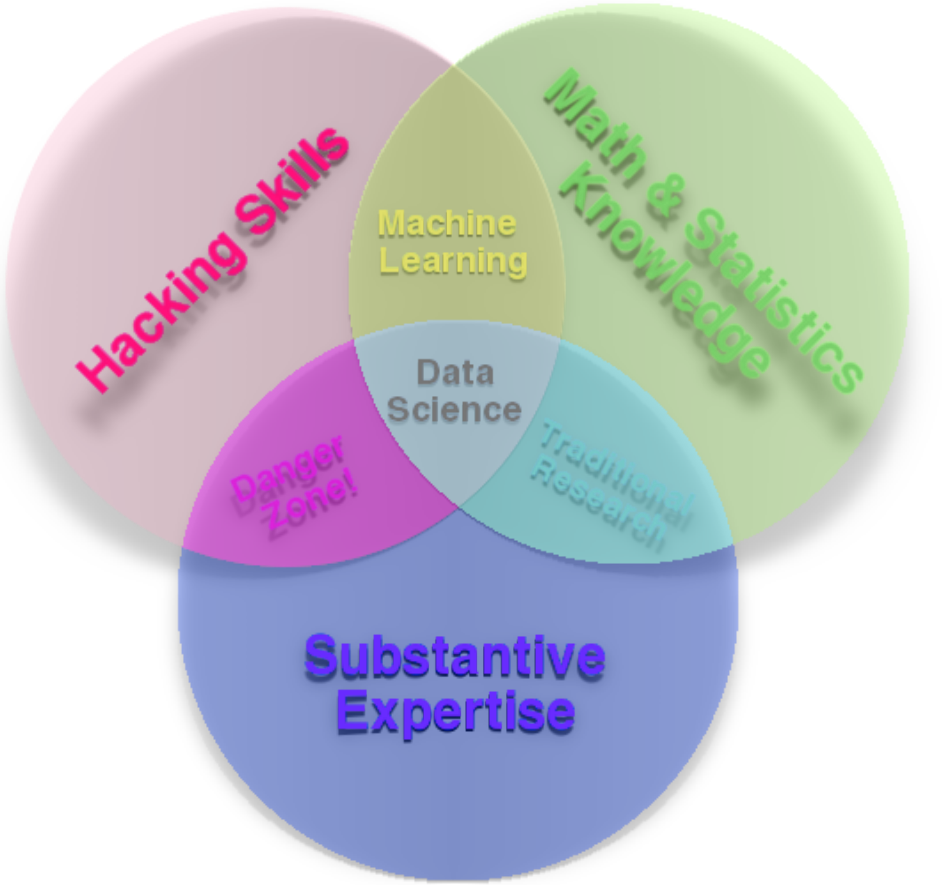
Technology is the collection of tools, including machinery, modifications, arrangements and procedures used by humans.

Big Data is a key **technology** to process massive amounts of data (f.e. to count items).

Methodology is the systematic, theoretical analysis of the methods applied to a field of study.

Data Science is a **methodology** to define what we want to do with data, how do we evaluate our actions, what decisions can be grounded on data, how do we combine evidences from several sources, etc.

Data Science



Drew Conway's Data Science Venn Diagram

eurecat
Centre Tecnològic de Catalunya

- **Data Science** is **not** a **science**

but a methodology based on multidisciplinar knowledge.

- *Currently, most company decisions are based on intuition and best practices. The alternative is to integrate data-based knowledge in the decision process.*

- Data Science is a new data processing model focused on turning data into actions.

Data Science

Steps:

- Ask a question.
- Get the data. They can be heterogeneous and non structured.
- Data Processing (cleaning, munging, etc.).
- Data Analysis (computer science, linguistics, economy, sociology, etc.).
- Take a decision and act.

Data Science

Data Science

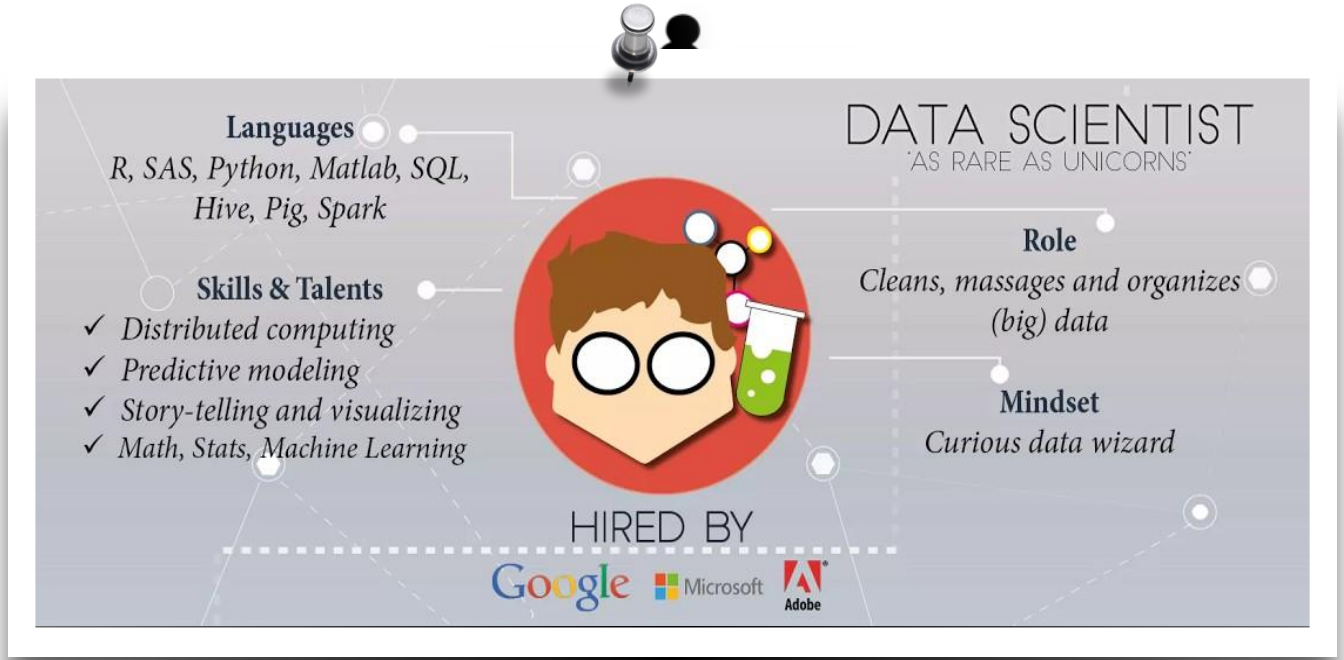
eurecat
Centre Tecnològic de Catalunya

What are the limits of data science?

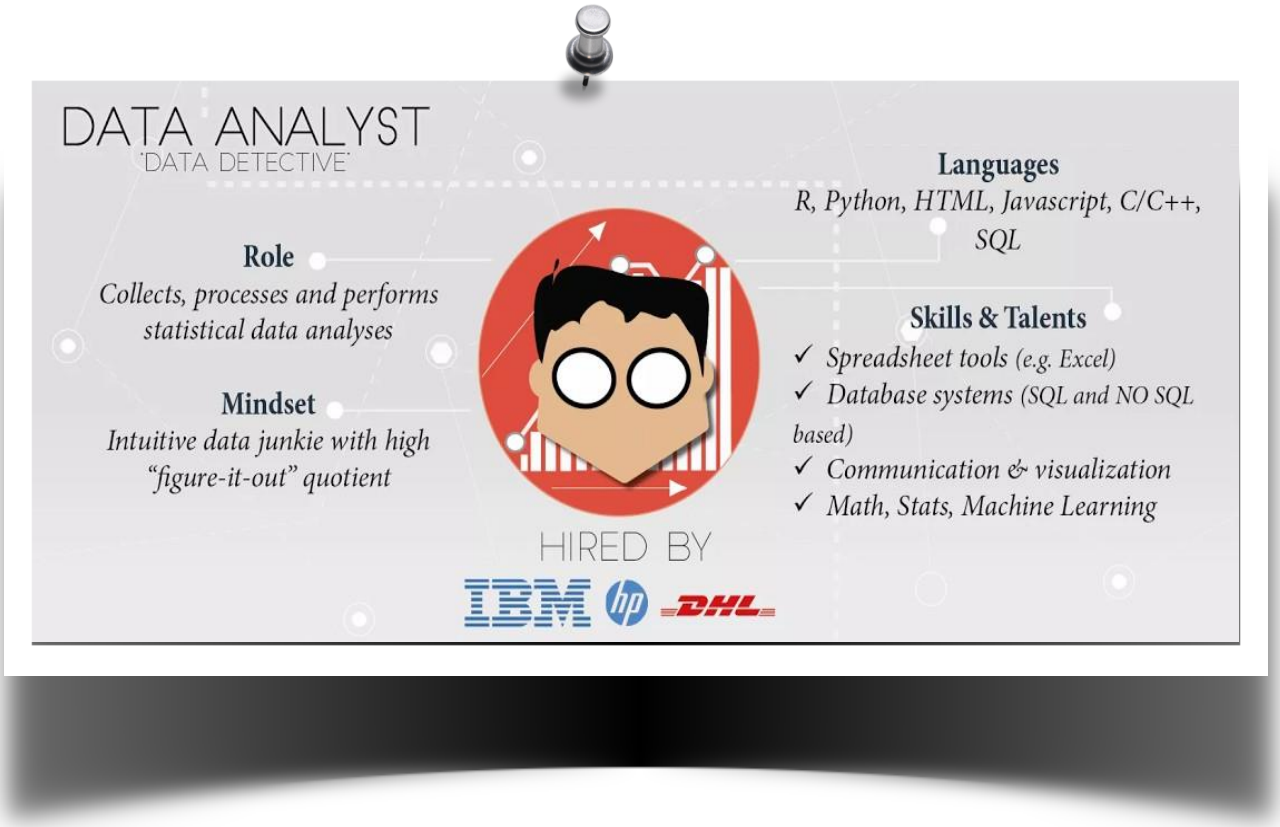
- Data science is a tool to inform, not to explain.
- Data science cannot substitute intuition or creativity.

If I had asked people what they wanted, they would have said faster horses.
Henry Ford.

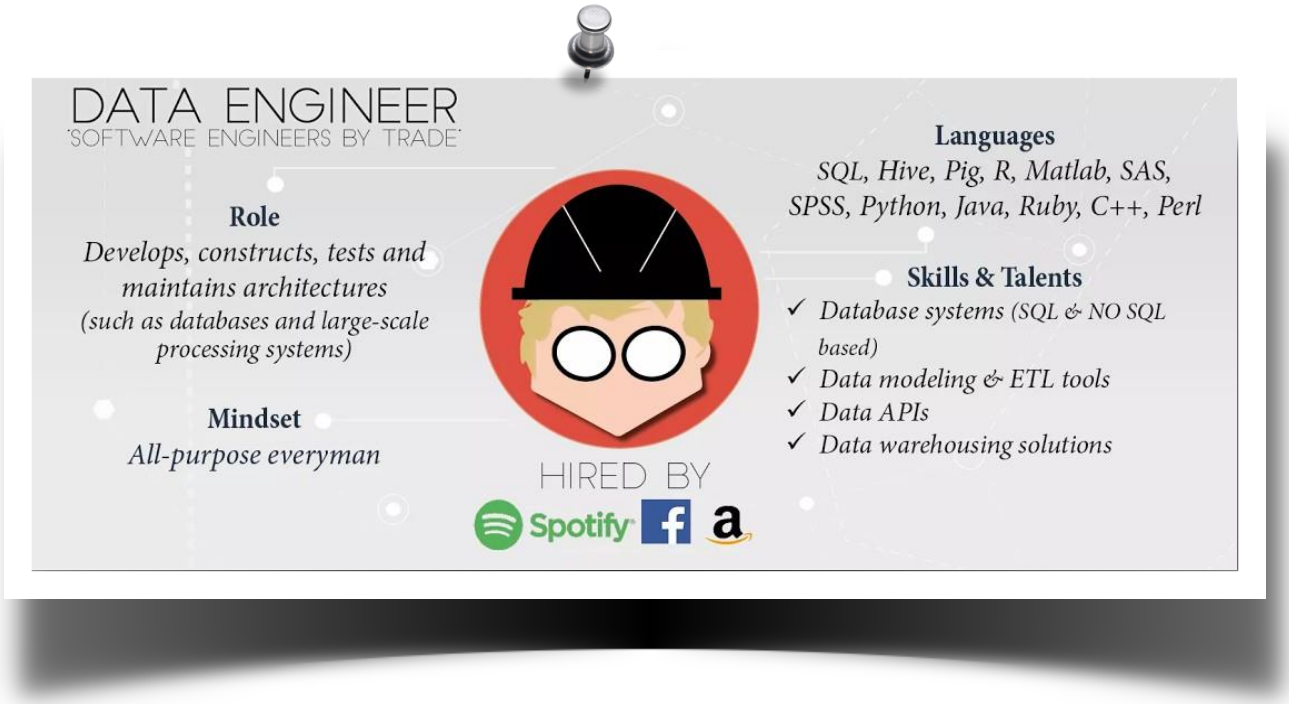
Big Data Roles



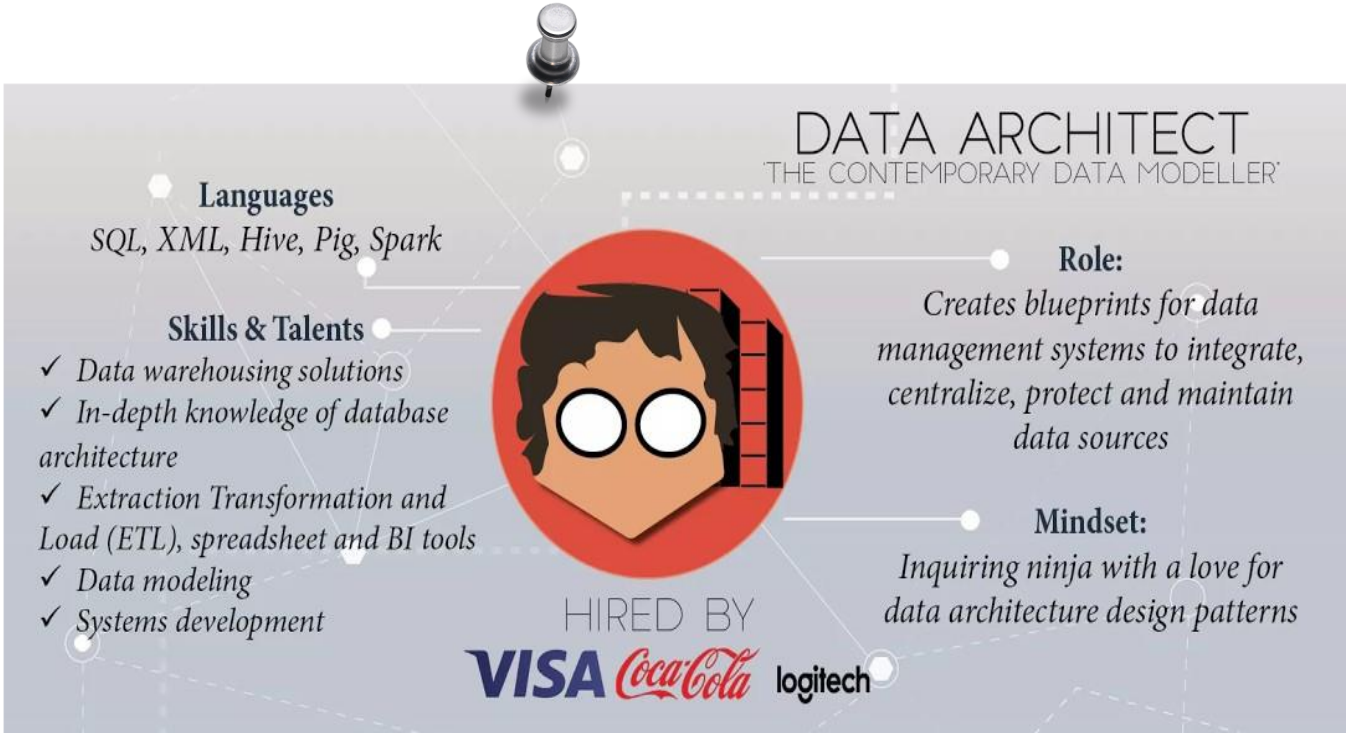
Big Data Roles



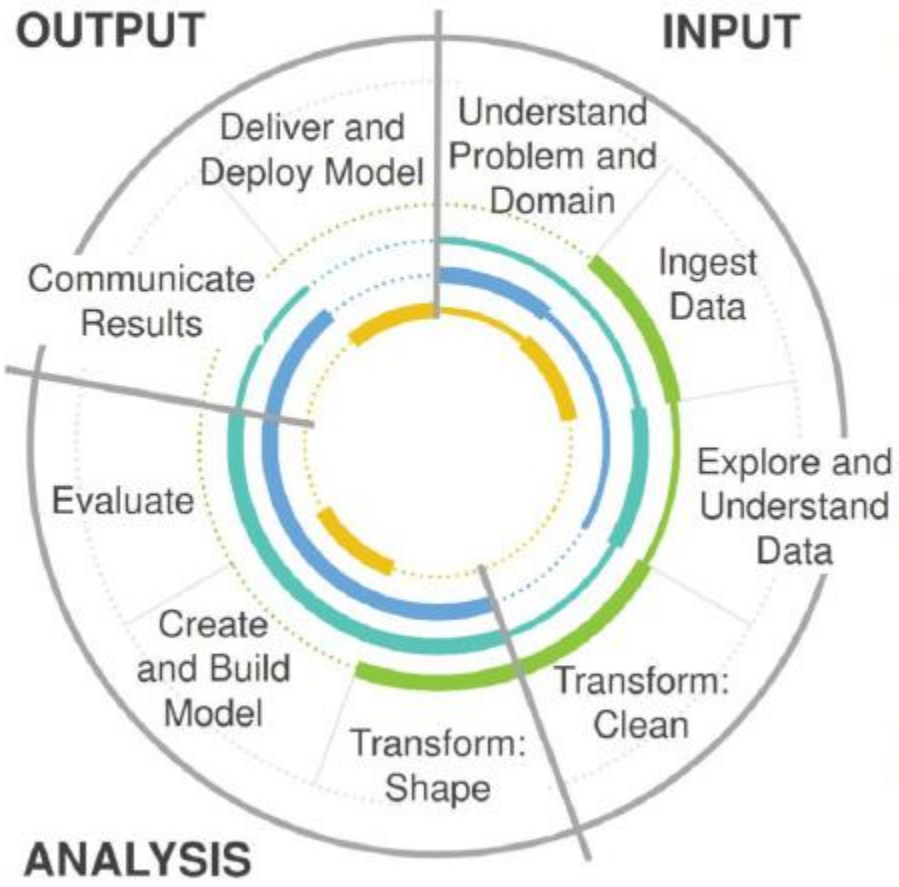
Big Data Roles



Big Data Roles



Big Data Roles



Data Engineer
Architects how data is organized & ensures operability

Data Scientist
Gets deep into the data to draw hidden insights for the business

Business Analyst
Works with data to apply insights to the business strategy

App Developer
Plugs into data and models & writes code to build apps

Big Data Roles

eurecat
Centre Tecnològic de Catalunya

The CDO role

Responsibilities of the Chief Data Officer – In Brief

- **Champion and safeguard enterprise data across the organization.** As many businesses have learned, and Debra Logan of Gartner points out, "CIOs...do not own the data. When retiring an asset, have you been able to get a straight answer from your business on how long to keep the data? They aren't making decisions, and you can't make the decisions about the data." Someone needs to manage data as the critical asset that it is.
- **Develop and oversee a corporate-wide data governance program.** To improve data quality, reduce risk and comply with all relevant regulations, teams must be created to implement data governance policies. In an organization with a Chief Data Officer, there should be no debate about "who owns the data."
- **Drive information and analytics strategy – from a business perspective.** The CIO remains ultimately responsible for information strategy from a technical or systems perspective, but the chief data officer should articulate KPIs and metrics that should be tracked by and reported on by the systems that the organization implements.

Big Data Roles

The CDO role

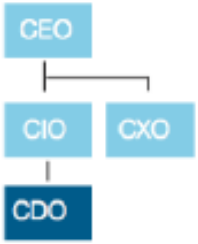
1. Direct report to CEO

34%

- ☑ Raise data to CEO's agenda for full empowerment
- ☑ Ensure data ownership independence from individual business units (BUs).
- ☒ Risk power issues between CIO and CDO

2. Integration within IT

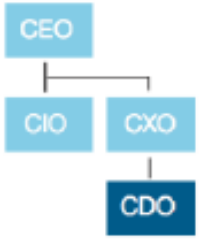
31%



- ☑ Leverage data as an operational enabler for business decisions
- ☑ Easily coordinate data management, architecture and IT
- ☒ Risk loss of business relevance for data

3. Report to functional area

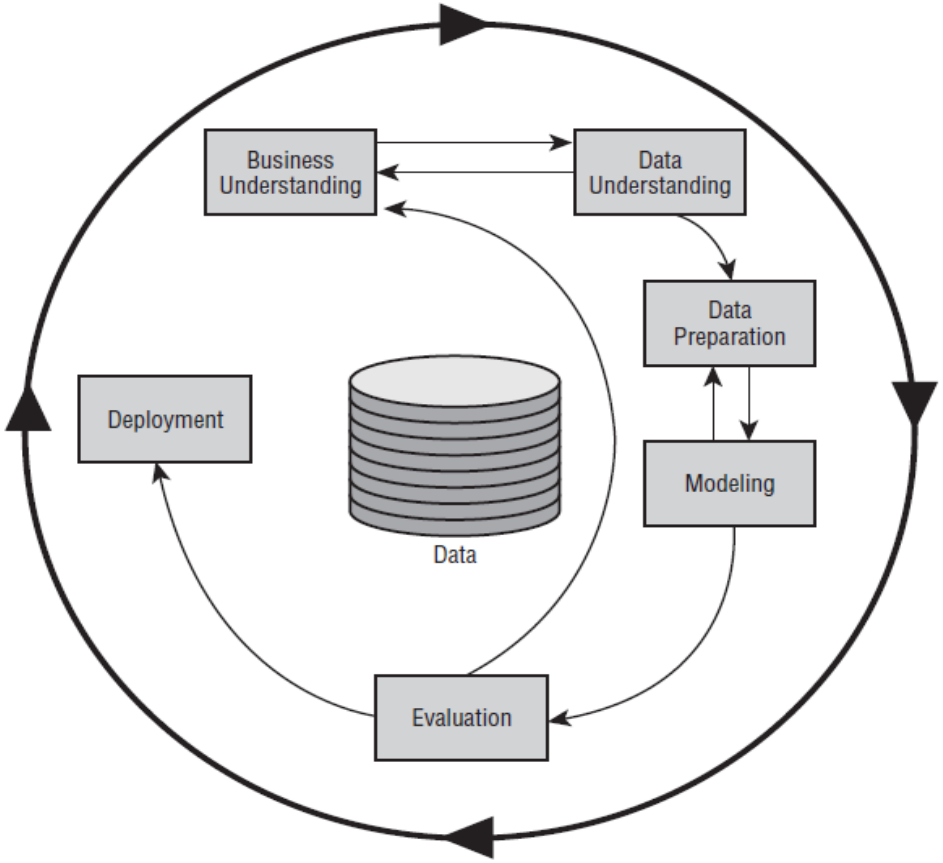
34%



- ☑ Provide data independence from IT
- ☑ Enable cross-BU functions for data management
- ☒ Risk diminished operational effectiveness caused by loss of IT contact

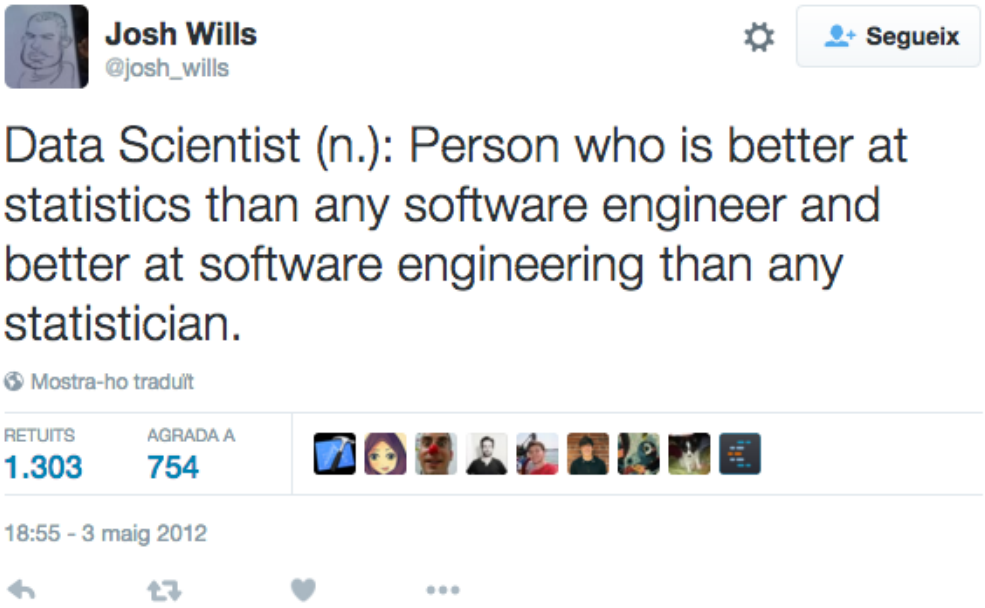
eurecat
Centre Tecnològic de Catalunya

Metodologies: Cross Industry Standard Process



eurecat
Centre Tecnològic de Catalunya

Metodologies



eurecat
Centre Tecnològic de Catalunya

02

The “Data Science” Toolbox

37

eurecat
Centre Tecnològic de Catalunya

Maths and Statistics

- Descriptive statistics
- Linear Algebra
- Numerical analysis
- Optimization
- Bayesian probability models

eurecat
Centre Tecnològic de Catalunya

Programming skills

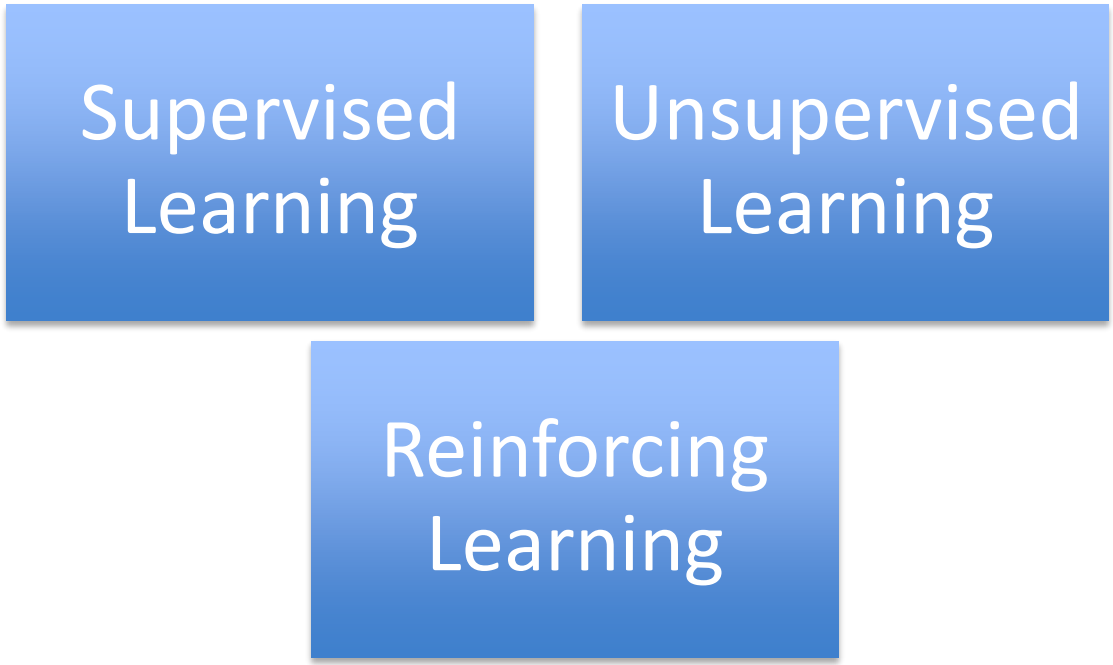
- Algorithm prototyping
- Programming languages for prototyping
  - ✓ R
  - ✓ Python
  - ✓ Matlab
  - ✓ Julia
  - ✓ Java
  - ✓ Scala
- Big Data Tools: Hadoop, Spark, Amazon WS, Kafka, etc.

eurecat
Centre Tecnològic de Catalunya

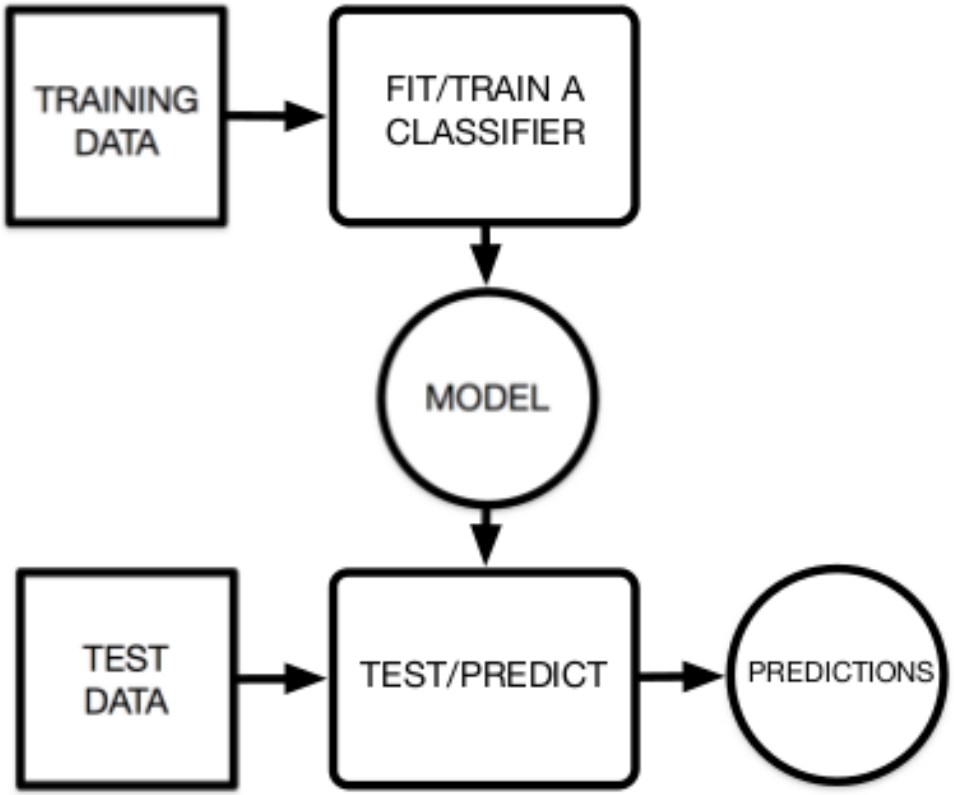
Techniques

- Classification and class probability
- Regression
- Similarity matching
- Clustering
- Co-ocurrence grouping
- Profiling
- Data reduction
- Casual modeling
- A/B testing

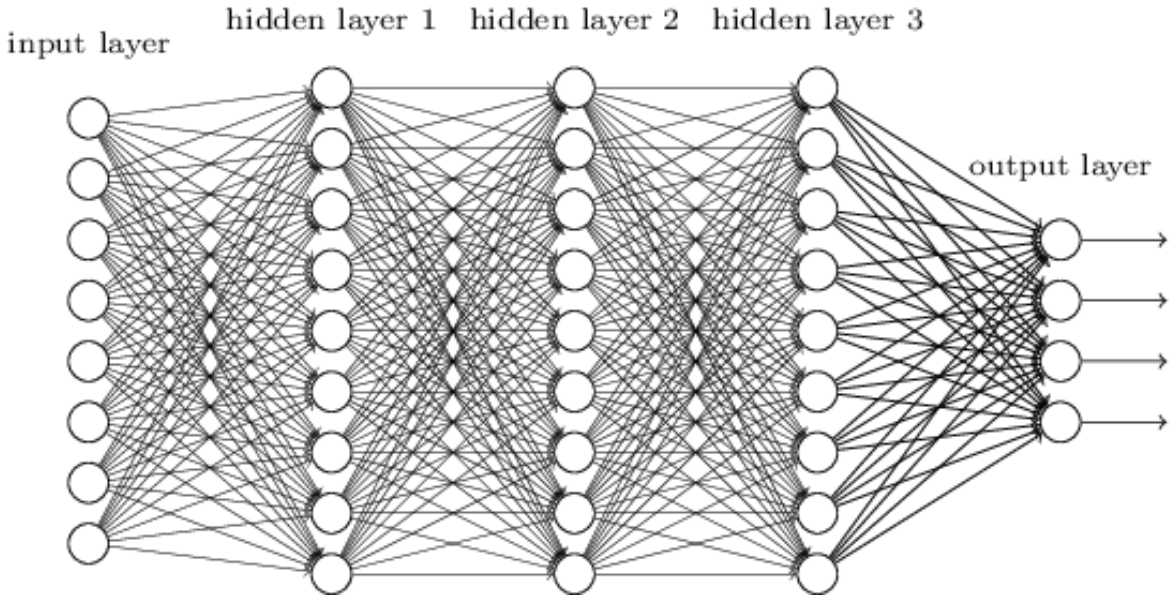
Machine learning



Machine learning workflow



And…what about deep learning?



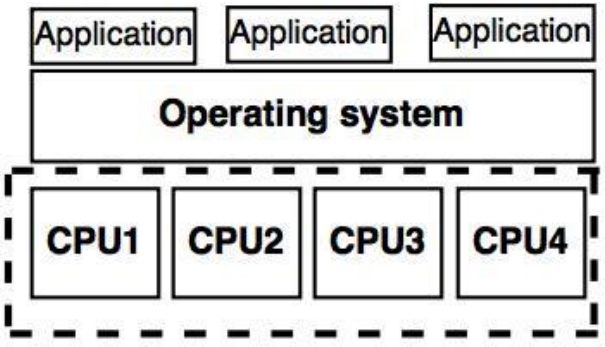
eurecat
Centre Tecnològic de Catalunya

03

The “Big Data toolbox”

44

Distributed computin and real time

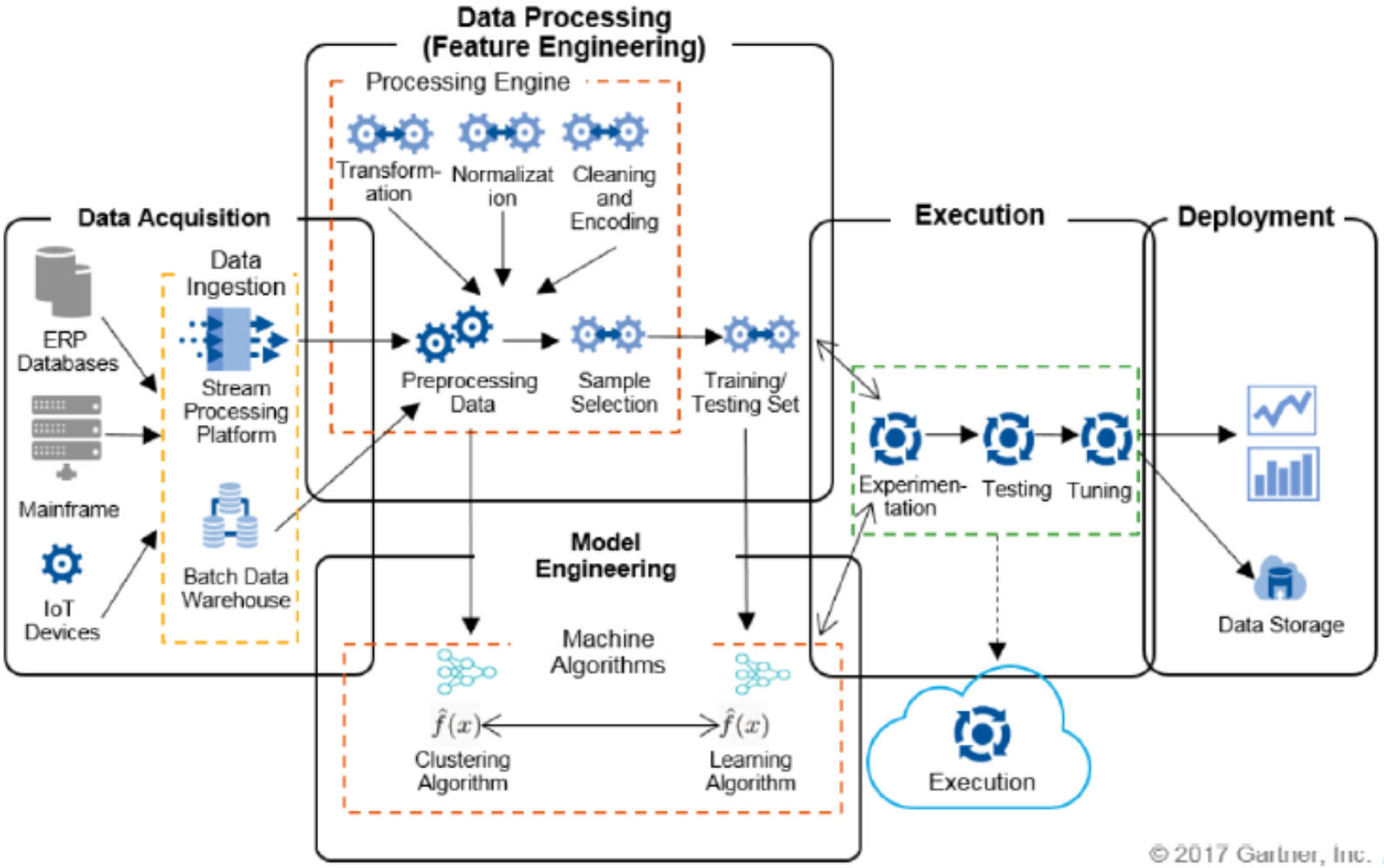


Virtualization on multicore processor

Batch

Stream

A Big Data architecture

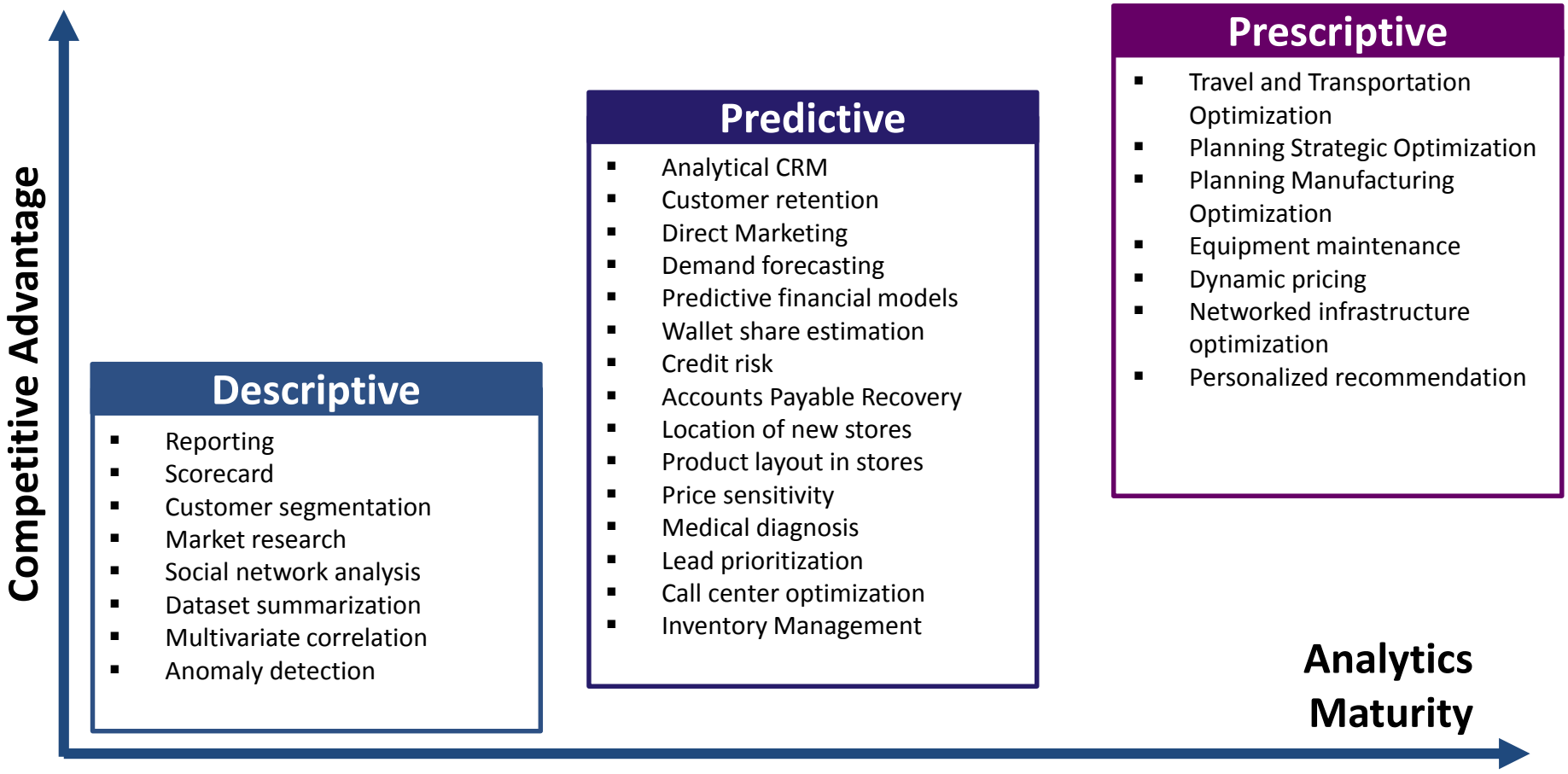


04

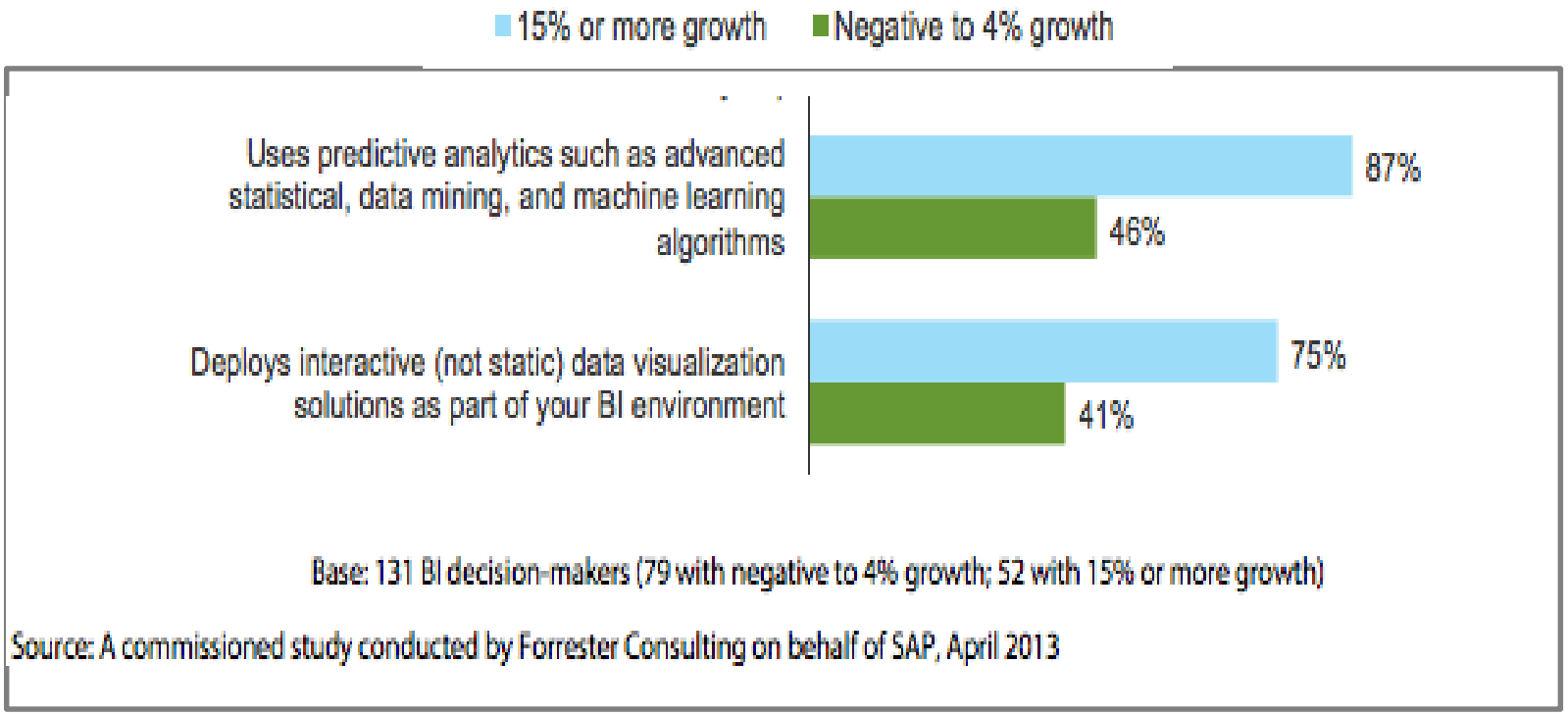
Market and industry

47

Data Analytics Capabilities



El valor per a les organitzacions



El valor per a les organitzacions

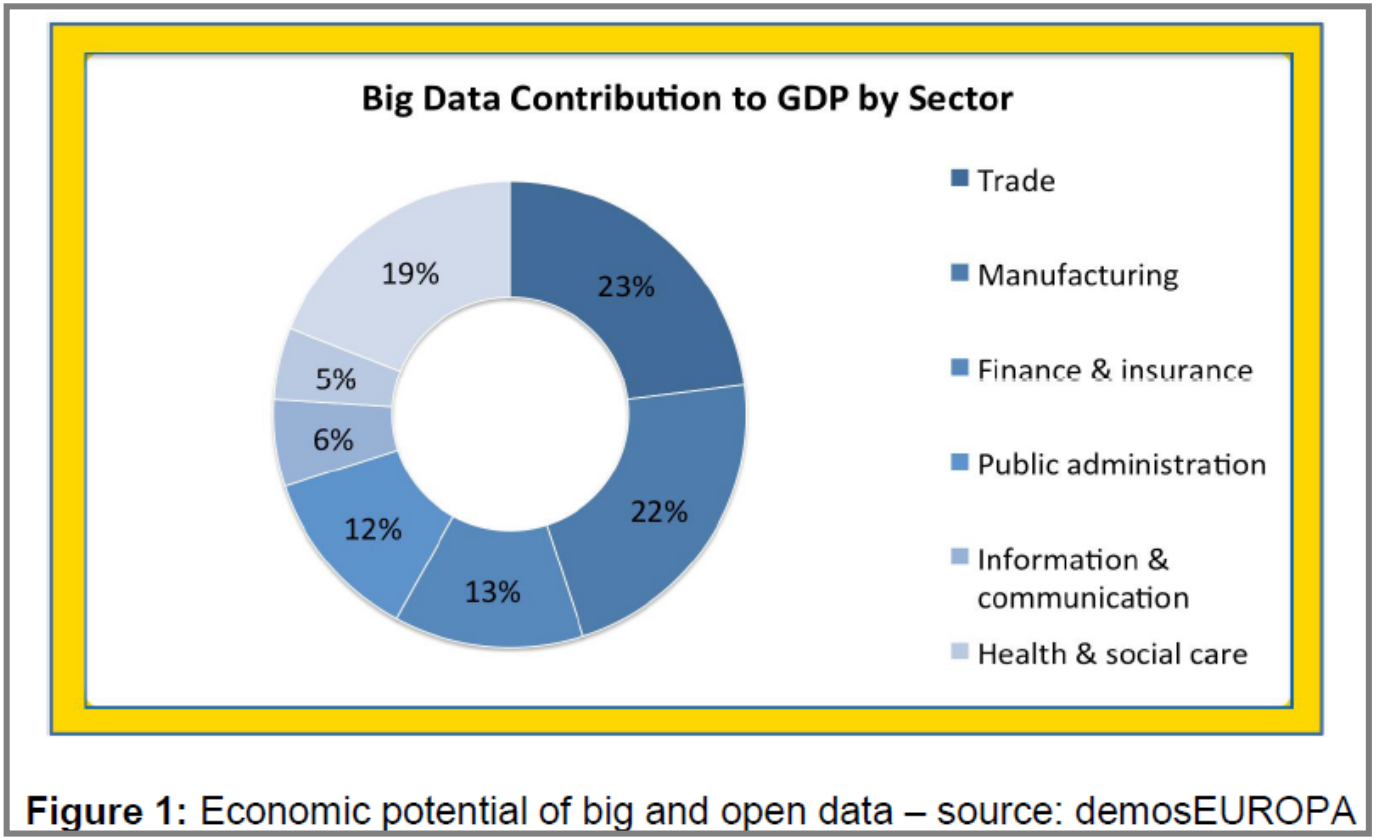


Figure 1: Economic potential of big and open data – source: demosEUROPA

Fields of application

Operational Efficiency

Use data to:

- Increase level of transparency
- Optimize resource consumption
- Improve process quality and performance

Customer Experience

Exploit data to:

- Increase customer loyalty and retention
- Perform precise customer segmentation and targeting
- Optimize customer interaction and service

New Business Models

Capitalize on data by:

- Expanding revenue streams from existing products
- Creating new revenue streams from entirely new (data) products

Big Data Challenges

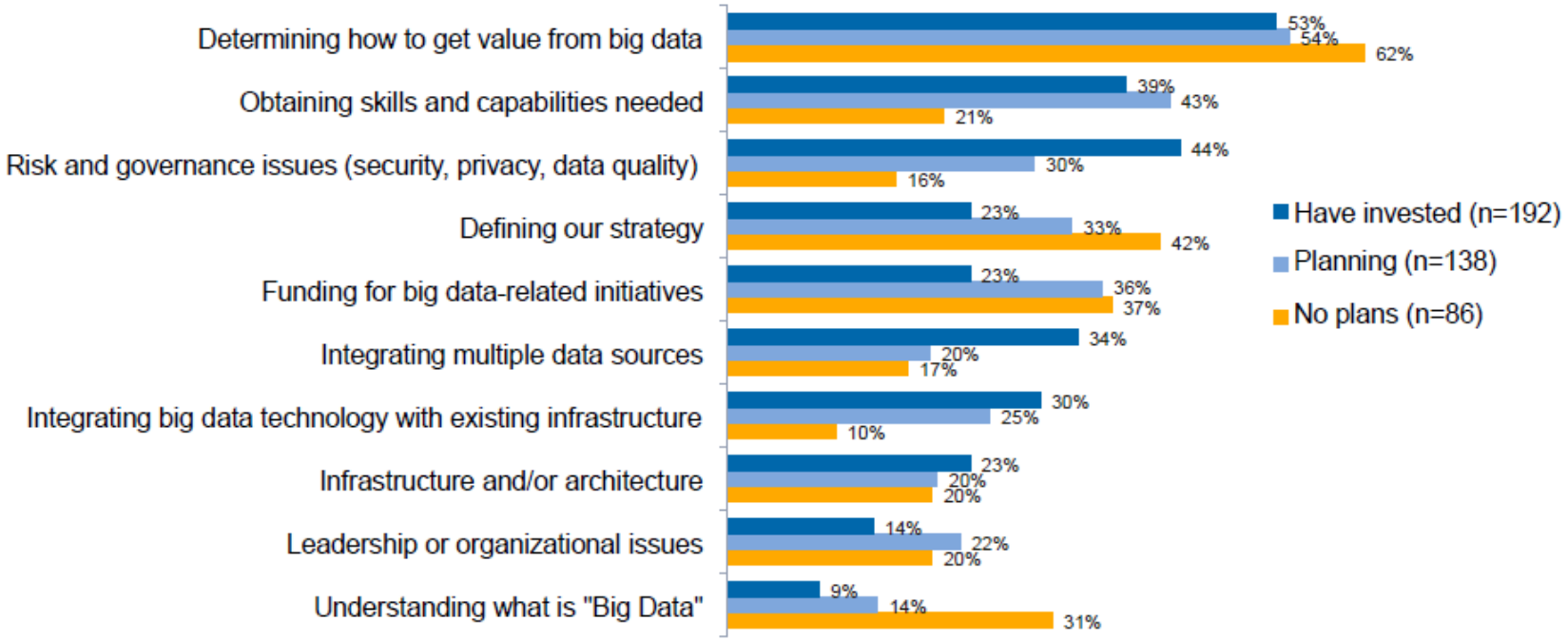


Figure 10. Percentage of Surveyed Customers Reporting Active Use, by Company Size, 2016

			Mobile	Cloud		Big Data Storage				Big Data Analytics			Data Analysis Techniques										
Company Size Segment	Employee Range	Number of References	Mobile BI Deployment	Cloud BI Deployment	Cloud Data Warehouse	Columnar/In-Memory Database	High-Capacity Data Warehouse	NoSQL Data Store	Hadoop Distribution	Search-Based Indexes	Complex-Event Processing	Advanced Analytics in MapReduce	Data Mining/Predictive Analysis	Log Data Analytics	Geospatial and Location Intelligence	Simulation/Optimization	Machine or Sensor Analytics	Link/Network/Graph Analytics	Text Analytics	Social Media Monitoring	Sentiment Analysis	Video Analytics	Speech Analytics
Small Business	1 to 99 employees	264	18%	54%	28%	31%	27%	10%	5%	17%	14%	7%	40%	26%	32%	22%	12%	15%	16%	12%	12%	3%	1%
Midsize Business	100 to 249 employees	183	18%	47%	27%	33%	25%	13%	7%	19%	6%	3%	35%	21%	24%	23%	8%	14%	14%	11%	9%	0%	1%
	250 to 499 employees	156	15%	39%	19%	28%	25%	6%	6%	15%	6%	8%	36%	21%	23%	17%	7%	11%	10%	5%	5%	0%	0%
	500 to 999 employees	173	23%	36%	15%	30%	23%	7%	8%	13%	4%	6%	27%	13%	15%	13%	15%	14%	8%	5%	4%	2%	1%
Large Business	1,000 to 2,499 employees	243	20%	30%	14%	30%	20%	2%	5%	12%	4%	7%	31%	19%	22%	17%	6%	5%	7%	7%	6%	1%	0%
	2,500 to 4,999 employees	189	18%	22%	10%	28%	26%	11%	8%	11%	4%	9%	26%	29%	19%	16%	14%	13%	8%	5%	3%	1%	2%
	5,000 to 9,999 employees	175	25%	27%	14%	34%	20%	8%	8%	20%	8%	8%	36%	25%	26%	23%	15%	14%	12%	16%	13%	2%	2%
	10,000 to 24,999 employees	231	17%	25%	13%	38%	30%	11%	13%	23%	14%	16%	35%	29%	25%	23%	14%	18%	15%	13%	11%	2%	3%
	25,000 to 49,999 employees	120	22%	24%	12%	49%	50%	13%	16%	26%	12%	9%	54%	30%	36%	32%	18%	23%	22%	15%	23%	5%	6%
	50,000 or more employees	276	23%	27%	11%	41%	36%	16%	26%	23%	15%	18%	42%	36%	38%	30%	20%	14%	18%	15%	17%	4%	3%
Overall Survey		**2,244**	**20%**	**34%**	**16%**	**33%**	**27%**	**9%**	**10%**	**18%**	**9%**	**10%**	**35%**	**25%**	**26%**	**21%**	**13%**	**13%**	**13%**	**10%**	**10%**	**2%**	**2%**

BI = business intelligence

Source: Gartner (October 2016)

Figure 6. Percentage of Surveyed Customers Reporting Active Use, by Industry, 2016

		Mobile	Cloud		Big Data Storage				Big Data Analytics			Data Analysis Techniques										
Industry	Number of References	Mobile BI Deployment	Cloud BI Deployment	Cloud Data Warehouse	Columnar/In-Memory Database	High-Capacity Data Warehouse	NoSQL Data Store	Hadoop Distribution	Search-Based Indexes	Complex-Event Processing	Advanced Analytics in MapReduce	Data Mining/Predictive Analysis	Log Data Analytics	Geospatial and Location Intelligence	Simulation/Optimization	Machine or Sensor Analytics	Link/Network/Graph Analytics	Text Analytics	Social Media Monitoring	Sentiment Analysis	Video Analytics	Speech Analytics
Banking	193	14%	20%	14%	38%	35%	13%	18%	20%	15%	18%	47%	25%	26%	31%	11%	11%	18%	11%	14%	1%	2%
Communications	82	18%	25%	18%	39%	42%	24%	17%	28%	11%	13%	45%	35%	35%	26%	21%	20%	17%	16%	19%	5%	5%
Education	75	12%	35%	14%	18%	15%	0%	0%	8%	3%	2%	25%	21%	23%	11%	8%	12%	5%	4%	8%	0%	0%
Government	147	17%	19%	8%	18%	19%	1%	1%	15%	7%	8%	25%	24%	26%	16%	7%	16%	11%	6%	4%	1%	1%
Healthcare	125	17%	21%	3%	37%	23%	2%	3%	17%	0%	2%	27%	27%	27%	13%	7%	9%	12%	5%	6%	2%	4%
Insurance	137	22%	23%	10%	26%	26%	14%	9%	19%	8%	13%	45%	18%	28%	17%	3%	10%	9%	5%	6%	0%	0%
Manufacturing and Natural Resources	377	24%	23%	13%	34%	24%	8%	6%	15%	5%	4%	30%	23%	16%	20%	17%	9%	11%	6%	6%	2%	1%
Media	82	29%	56%	33%	52%	48%	23%	22%	30%	15%	13%	48%	46%	44%	33%	20%	27%	18%	42%	18%	8%	4%
Other	204	14%	36%	10%	26%	19%	9%	9%	11%	5%	7%	22%	14%	21%	15%	8%	13%	11%	11%	9%	3%	3%
Retail	217	20%	32%	16%	40%	33%	9%	8%	15%	10%	9%	35%	20%	21%	18%	9%	13%	9%	8%	8%	0%	0%
Services	454	20%	55%	32%	38%	30%	12%	15%	23%	12%	15%	39%	30%	37%	26%	19%	17%	17%	16%	16%	3%	2%
Transportation	72	15%	28%	6%	36%	24%	2%	9%	20%	12%	5%	36%	35%	37%	28%	18%	8%	4%	4%	9%	0%	2%
Utilities	37	40%	24%	7%	19%	7%	13%	13%	0%	8%	7%	47%	29%	25%	29%	31%	14%	18%	13%	0%	0%	0%
Wholesale Trade	42	23%	32%	7%	23%	23%	0%	0%	18%	8%	4%	19%	11%	0%	8%	4%	7%	12%	4%	11%	0%	0%
Overall Survey	**2,244**	**20%**	**34%**	**16%**	**33%**	**27%**	**9%**	**10%**	**18%**	**9%**	**10%**	**35%**	**25%**	**26%**	**21%**	**13%**	**13%**	**13%**	**10%**	**10%**	**2%**	**2%**

BI = business intelligence
Data for industries with fewer than 50 references should be considered merely directional.

Source: Gartner (October 2016)

Business priorities

	Manu & N. Res.	Media/ Comm	Svcs	Gov.	Edu	Retail	Banking	Insur-ance	Health-care	Trans-portation	Utilities
Enhanced customer experience	52%	78%	66%	43%	76%	83%	77%	77%	73%	69%	44%
Process efficiency	45%	33%	35%	49%	65%	43%	41%	50%	73%	69%	78%
More targeted marketing	43%	89%	53%	17%	41%	78%	66%	58%	-	38%	17%
Cost reduction	42%	33%	35%	37%	35%	30%	41%	31%	45%	56%	61%
Improved risk management	14%	22%	29%	29%	35%	22%	52%	58%	55%	31%	61%
New products	23%	67%	37%	14%	24%	35%	27%	50%	-	19%	33%
Developing information products	26%	33%	44%	31%	12%	22%	23%	19%	9%	19%	11%
Enhanced security capabilities	17%	22%	21%	34%	29%	13%	27%	27%	9%	19%	28%
Regulatory compliance	11%	22%	18%	23%	18%	9%	25%	23%	27%	31%	44%
n=	65	9	62	35	17	23	44	26	11	16	18

Multiple responses allowed

Definition

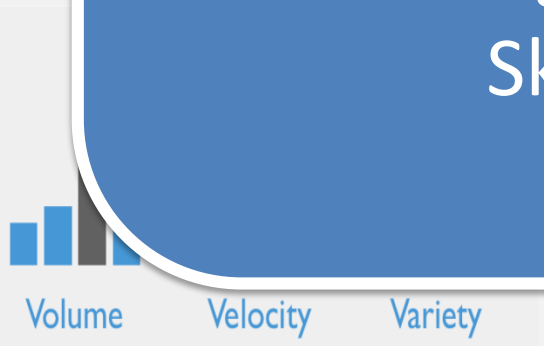
eurecat
Centre Tecnològic de Catalunya

Today however

LACK OF

Data Strategy
Central Repository
Right Data Quality
Skills and Culture

"Big dat
variety
effectiv
processi
making.



What is “Big” Data for me?

eurecat
Centre Tecnològic de Catalunya

Intuitive and interactive user interfaces

Base technologies

Enhanced Insight

Advan

Big Data Infrastructures

Improved Decision Making

Process Automation

Enhanced Customer Experience

Gartner defines customer experience as the customer's perceptions and related feelings caused by the one-off and cumulative effect of interactions with a supplier's employees, systems, channels or products.

Recommender Systems

Personalized recommendations and multi-level reward programs based on historical data like purchase preferences and customer behavior. Dades de novestendències i revenues

Sentiment Analysis

Natural language processing of social media streams, call center records, product reviews, etc., for customer feedback and market insights.
Promising advances through Deep Learning approach

Customer 360

Real time decisioning can increase customer value by determining the Best Next Action across all channels of communication with customers. Models based on machine learning techniques to predict customer behavior and manage multi-channel interactions

By 2017, 89% of marketers expect customer experience to be their primary differentiator. Gartner

The Effectiveness of Personalized Product Recommendations

MarketingSherpa Study, 1.5 billion shopping sessions, 2015:

The recommendations used a variety of different common phrases on a product page, home page, shopping cart, category page or site wide. The actual product recommendations were dynamic and personalized based on visitor data, behavior, and history.

- On the whole, **11.5% of the revenue** (whether from more volume or higher value of products) generated in the shopping sessions was attributable to **purchases from the product recommendations.**
- The companies that used the most common "**visitors who viewed this product also viewed"** on the product page had the highest success, **with a remarkable 68% of all revenue of those companies coming from the product recommendations.**
- The phrasing **"you might also like, "** correlating to **16% of that group's revenue to the recommendations.**
- The popular phrasing **"customers also bought"** on the cart page generated only **8% of revenues from recommendation sales.**

HMV, a British entertainment retailing company (music Retailer) realized that sending the same campaign message to all its customers is not appropriate anymore, as people start treating emails as spam and do not open them. The company uses **a recommendation system**, which analyses customer click streams and which products fits the customer's preferences. HMV sends out personalized recommendations, which increased the emails opening by over 70% on mobile phones, and PC mails by 50 %.

In 2013 **Item-to-Item collaborative filter:**
35% of all sales are estimated to be generated by the recommendation engine.
In May 2016, Amazon opened up **DSSTNE** as open source software so that the promise of deep learning can extend beyond speech and object recognition to other areas such as search and recommendations

After a long refinement process, Netflix finally released its first "global" recommendation engine in December, 2016. Netflix will invest 1 billion of the total 5 billion of its budget in recommendation and personalization.
Why?? Netflix estimates that only 20% of its subscriber video choices come from search, with the other 80% coming from recommendations

Success Cases: Social analytics

Social media has emerged as a platform where people reveal a lot about their personal tastes, preferences, likes and dislikes, whether they know it or not. The pages they like, or the posts they share, lead to a better understanding of a buyer's persona.

U.K. retail giant Argos uses social data to glean real-time insights on store performance, customer experience and brand reputation. "Using social insights, we can understand which stores are performing well for retail customers and identify areas for improvement," Argos' customer and digital insight manager James Finch said in 2015.

- **Predict what customers want before they ask for it.** Companies use social media information and cookie-based ad retargeting to put products or service recommendations into social feed and in front of the eyes of individual consumers most likely to buy.
- **Fix customer issues when they occur.** Analyzing consumers' posts or interacting with them on social media can help brands detect and solve their problems on a real-time basis. The resolution time is shorter because the data is right in front of the customer service representative. The individual attention helps increase the level of customer satisfaction and loyalty.
- **Discover and resolve customer pain points.** Big data allows companies to spot and address common or recurring customer complaints by collecting and analyzing numerous posts and tweets over a period of time.

Targeted Marketing

Targeted marketing is the process of identifying customers and promoting products and services via mediums that are likely to reach those potential customers.

MicroSegmentation

Breaking the market into very thin segments and then concentrating your marketing efforts on one or a few key segments

Next Best Offer

Tailored offers through online behavioral analysis and web analytics. Targeted campaigns using analytics to segment consumers, identify the most appropriate channels and achieve optimal ROI.

Predictive analytics in Marketing

Main steps:

1. Dynamic and ongoing segmentation of customers into small groups (**micro-segments**) who will likely behave similarly in response to marketing actions
2. **Behavior modeling** to predict how each micro-segment of customers will respond to each available marketing action
3. **Customer lifetime value** forecasting to predict the long-term impact of marketing actions on customers (not just the immediate-term results) at every point in the customer lifecycle
4. **Self-learning, closed-loop action optimization methodology** for testing, tracking and optimizing how marketing actions affect micro-segments of customers.

Examples

Next Best Action

Next product prediction

– Purchased, added to the cart, recommended, etc.

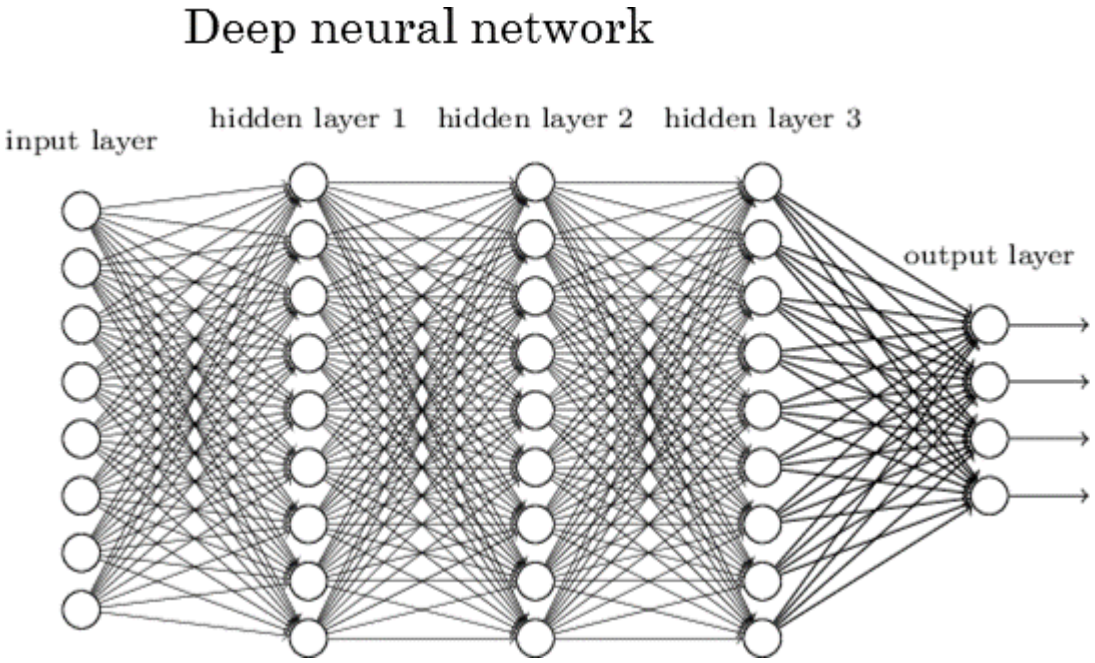
• Sequential input problem

– Sequence-based or item based

• **Recurrent Neural Networks (RNNs) can be used**

– Take into account time dependencies

– Raw data as opposed to time-based aggregated



Andrej Karpathy, "The Unreasonable Effectiveness of RNNs"

In 2017…

Why Deep Learning??

- ✓ GPU's can be used for training the networks
- ✓ No Feature-Engineering needed.
- ✓ Accuracies achieved are higher than those obtained with Machine Learning: DeepFace (DL application developed by Facebook) is more successful than humans in recognizing whether two images are of the same person or not .

DeepFace scoring a 97% success rate compared to expert humans with 96%

Examples

Customer Loyalty

Churn prediction is the task of identifying whether users are likely to stop using a service, product, or website.

- Churn Prediction model based on Machine Learning:

Decision Trees, SVM, Logistic Regression
Ensembles (Random Forests)
Boosting

Final method overall performance:

64 % accuracy on users who did churn
74% accuracy on users who did not churn

Probability Range Bins	% of users in bin that actually churn
0% – 10%	4.4%
10% – 20%	14.1%
20% – 30%	25.8%
30% – 40%	35.5%
40% – 50%	44.9%
50% – 60%	55%
60% – 70%	65.5%
70% – 80%	76.8%
80% – 90%	82.5%
90% – 100%	NaN

Optimize efforts: it is not worth trying to retain the 4.4% lower. We should focus only on 82.5% higher.

Look beyond just overall accuracy

Recent tests in churn predicition using Deep Learning show an overall accuracy higher than 78%.

Process efficiency and cost reduction

Main trends

Process efficiency is related to whether a company's processes are harnessing resources in the best way possible, whether those resources are financial, human, technological or physical. Process efficiency usually involves cost reduction.

Inventory Management and Demand Forecast

Every retailer wishes to have the right product at the right place at the right time. Predictive analytics technology will significantly improve demand forecast accuracy, and suggest better allocation and replenishment strategies.

Route Planning Optimization

more efficient transportation using C enabled big data telematics

Predicitive maintenance

Machine Learning provides a complementary approach to maintenance planning by analyzing significant data sets of individual machine performance and environment variables, identifying failure signatures and profiles, and providing an actionable prediction of failure for individual parts.

Examples *Optimal Shop location*

In 2007 and 2008, Starbucks' CEO Howard Schultz was forced to come out of retirement to close hundreds of stores, and rethink the company's strategic growth plan.

"This time around, Starbucks took a more disciplined, data-driven approach to store openings and used mapping software to easily analyze massive amounts of data about planned store openings. The software analyzed location-based data and demographics to determine the best place to open Starbucks stores without hurting sales at other Starbucks locations.

"The software is also helping to determine where the next 1,500-plus stores should be placed not only to help the company expand, but drive revenue for new store developments."

Data used:

- *Mobile data*
- *Demographic and income data (CENSUS)*
- *Geoinformation (OpenStreet Maps and Google)*

Big Data: Only for Big Players?

COMPANY
Mastercard

INDUSTRY
Finance

EMPLOYEES
67,000

TYPE
Behavioral Analytics

Purpose:

With 1.8billion customers, MasterCard is in the unique position of being able to analyze the behavior of customers in not only their own stores, but also thousands of other retailers. The company teamed up with Mu Sigma to collect and analyze data on shoppers' behavior, and provide the insights it finds to other retailers in benchmarking reports.

Big Data: Only for Big Players?

COMPANY
Coca-Cola Co.

INDUSTRY
Food

EMPLOYEES
146,200

TYPE
Market Basket Analysis

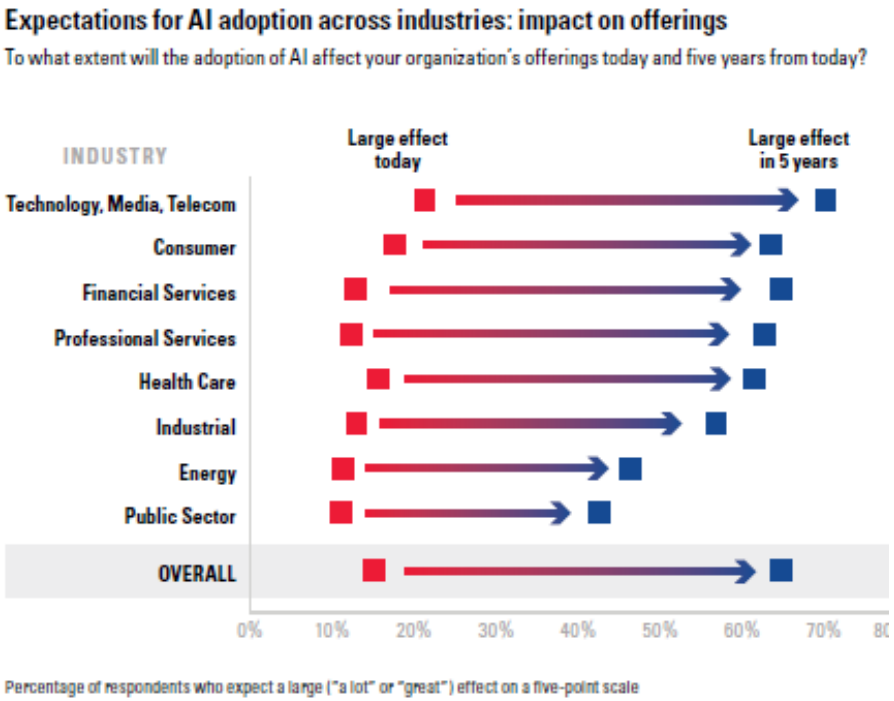
Purpose:

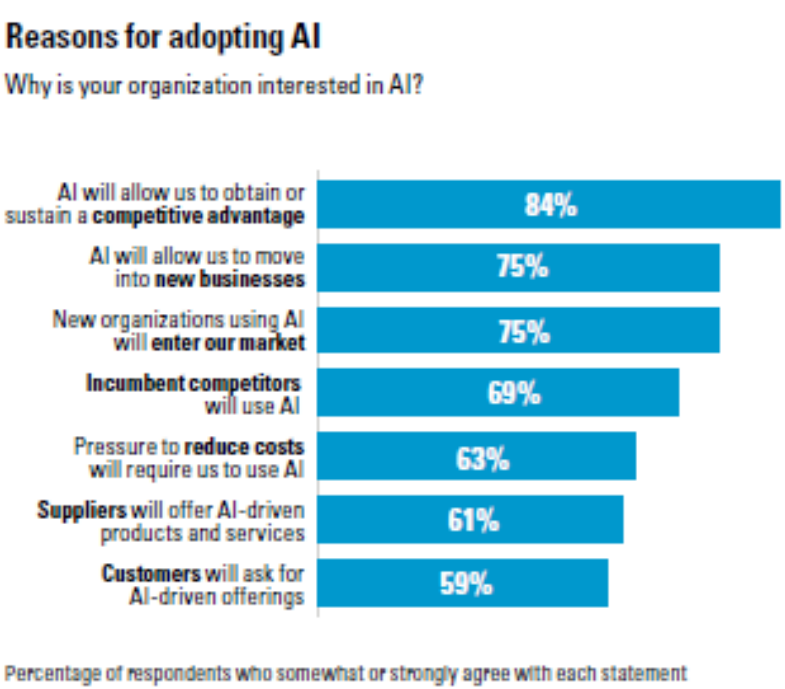
Coca-Cola uses an algorithm to ensure that its orange juice has a consistent taste throughout the year. The algorithm incorporates satellite imagery, crop yields, consumer preferences and details about the flavours that make up a particular fruit in order to determine how the juice should be blended.

05

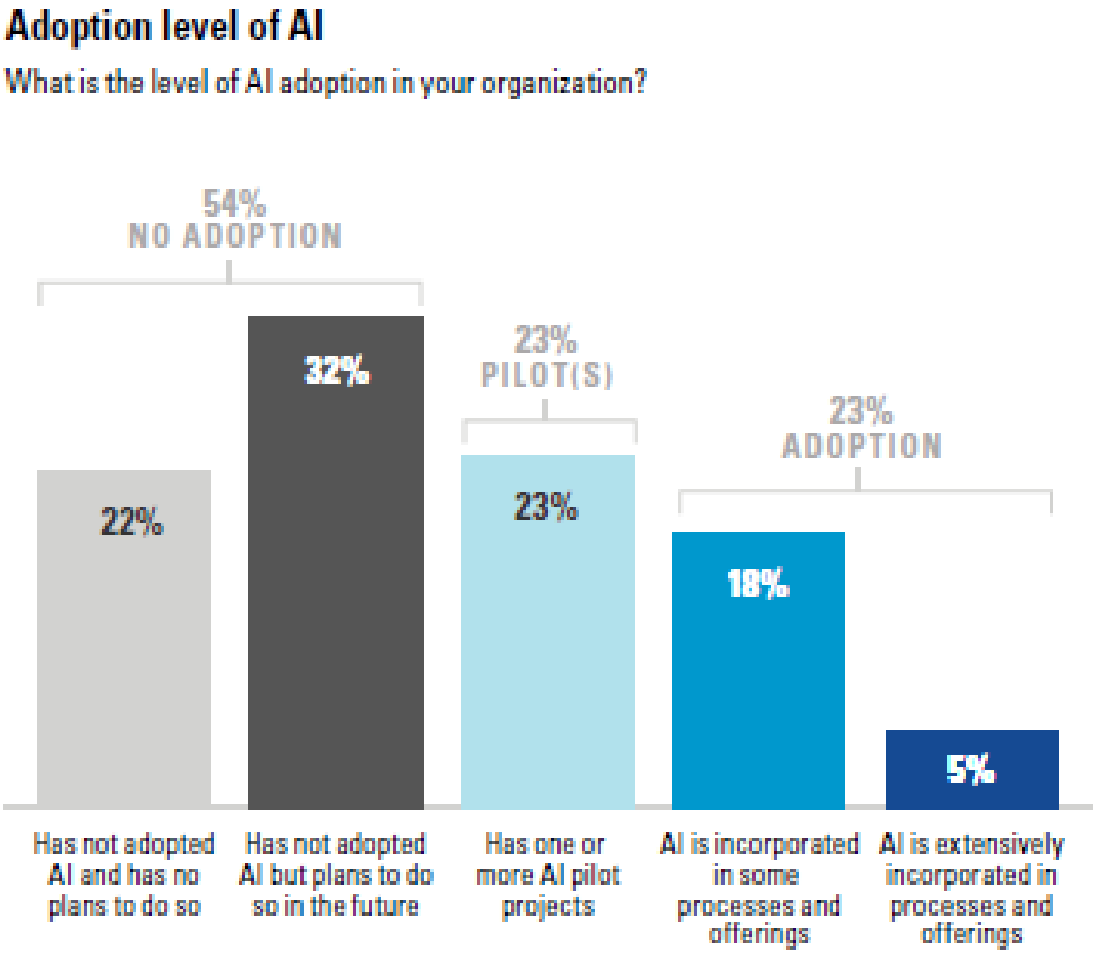
Artificial intelligence in the industry

Adoption of AI in business





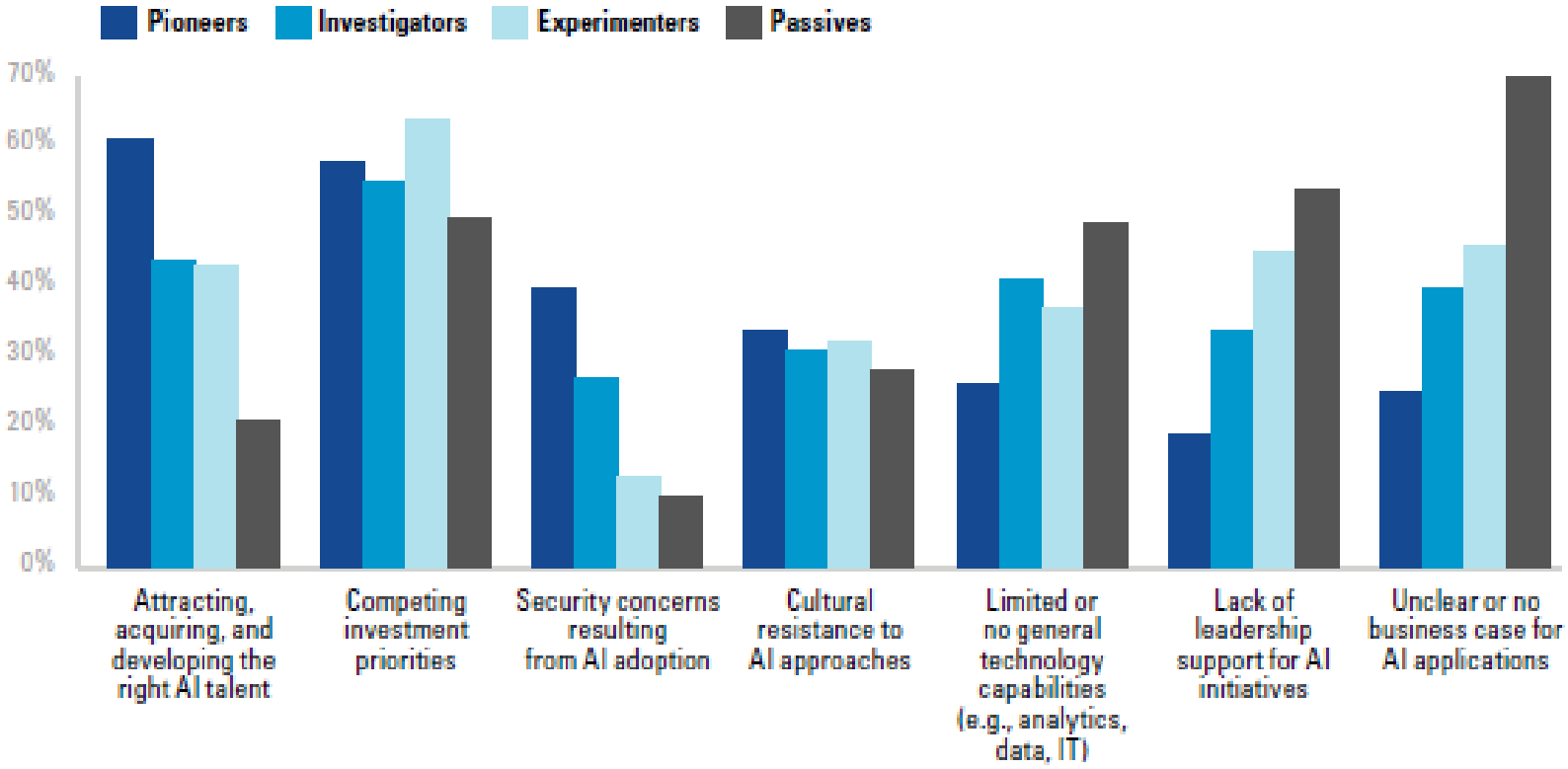
Adoption of AI in business



Adoption of AI in business

Barriers to AI adoption

What are the top three barriers to AI adoption in your organization?



Percentage of respondents ranking the selection as one of the top three barriers

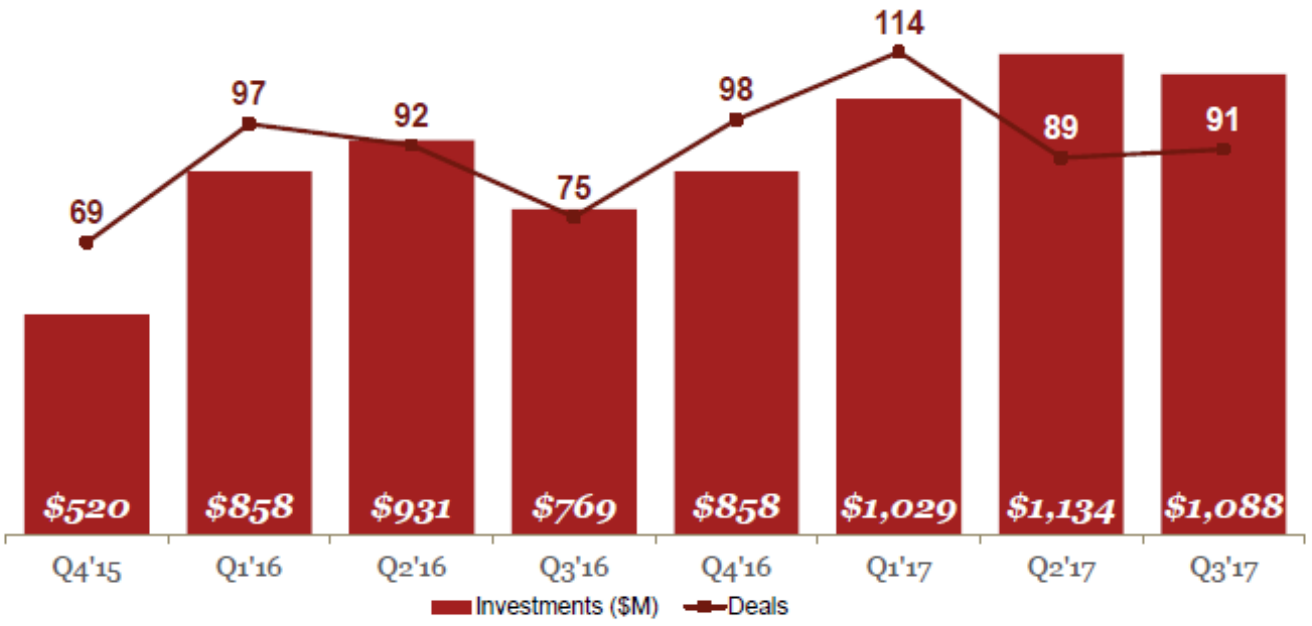
Is AI profitable?

eurecat
Centre Tecnològic de Catalunya

AI: US funding over time

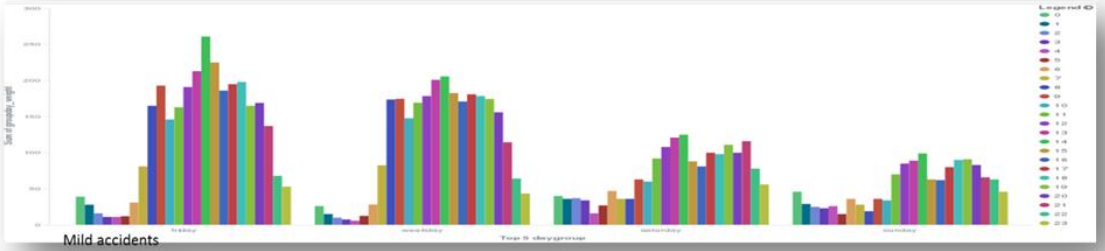
US artificial intelligence funding exceeds $1B for third quarter in a row

- Both deals and dollars to US AI companies showed continued momentum in Q3'17, with $1B invested across 91 deals. Quarterly funding was particularly healthy, nearly matching last quarter's high.
- The $1B quarter was led by companies such as NAUTO ($159M Series B), Indigo Agriculture ($156M Series D), and Cerebras Systems ($60M Series B).

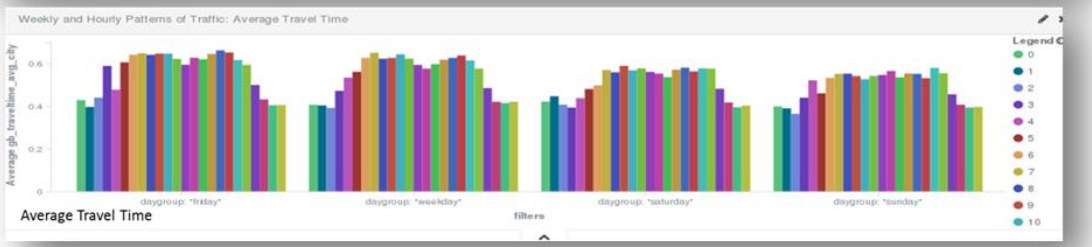


Few References Eurecat

Severe-mortal accidents within 150m of an education institution



Mild accidents



Few References Eurecat

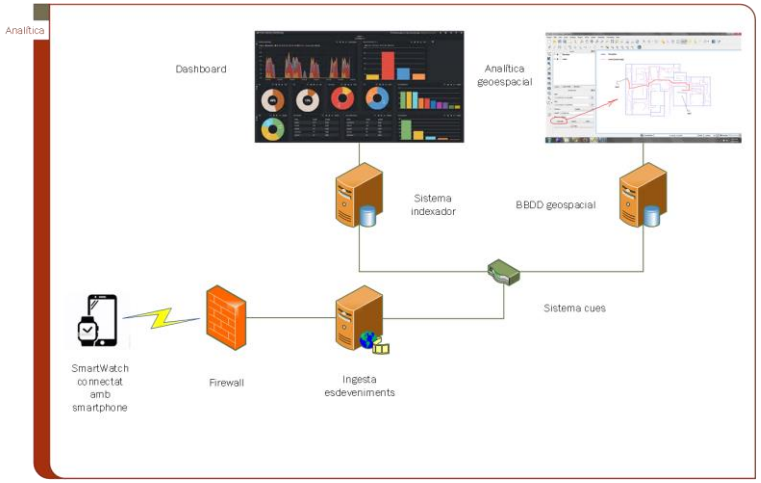
Recommendation system for smartwatch

OBJECTIVE:
Develop an application for smartwatch to personalize recommendations in the retail sector

DATA:
The interactions between the user and the application smartwatch, including indoor positioning data.

PROJECT DEVELOPMENT AND RESULTS:

- *Users:* A better user experience at the shopping time by means of the exploitation of the wearable sensors and App.
- *Business owners:* A better knowledge of the business thanks to the exploitation of the user data collected.



Few References Eurecat

Big Data platform & PoCs

OBJECTIVE:
The main goal of this project is to build the foundations of the ecommerce's Data Platform while proving some business value. This value is executed as a set of Proof of Concepts addressing business needs.

DATA:
Ecommerce data (campaigns, products, visits)

PROJECT DEVELOPMENT:
Technologies: AWS EMR, Sqoop, Apache Hive, Apache Spark and Apache Kafka.

RESULTS:
Deployment of the data platform (including a data lake) and development of the following Proofs of Concept:

- Product Ordering
- Product Recommendations
- Visit Recovery

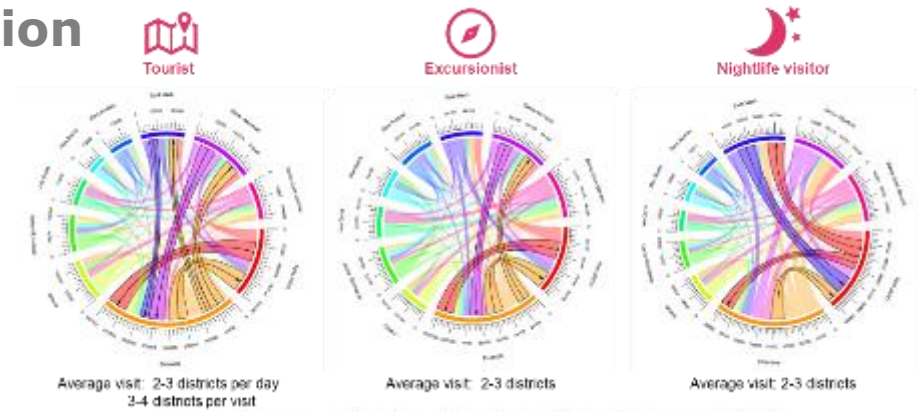
Confidential customer

Few References Eurecat

Big Data and Iot Tourism in action

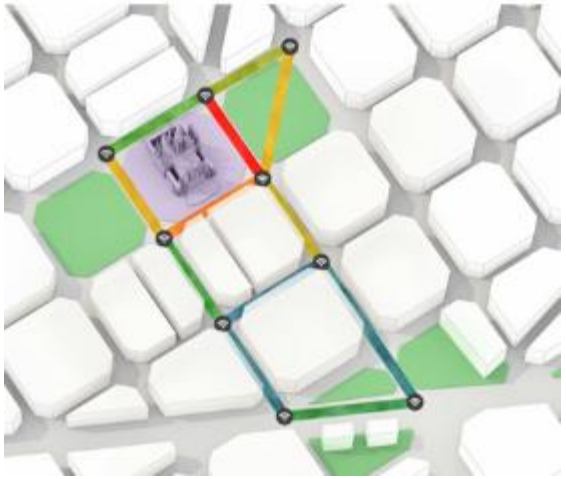
OBJECTIVE:

This initiative analyses concentrations, profile, behaviour and mobility flows of tourists in Barcelona and also at street level – in an area of high touristic interest: Sagrada Familia



DATA:

- Mobile phone data
- IoT devices (wifi, GSM, 3D cameras)



PROJECT DEVELOPMENT AND RESULTS:

Big Data and IoT devices are powerful tools to:

- Improve the knowledge of tourist activity
- Manage and optimize mobility flows and highly visited areas
- Improve decision making at the Tourism Area of the City Council

Few References Eurecat

WATENER

Benefits:

1. Energy cost reduction
2. Improvement of operational processes
3. Savings on company resources
4. Enhanced infrastructure planning

Karlsruhe, Germany

>300k inhabitants, 2.300Kwh/d

Thank you