

# “Artificial Intelligence and Big Data” 2018



# PROGRAM

## DAY 1 : Business Applications

- a. General Introduction : The New “Black Gold”
- b. Big Data for All !!!! - Examples of Real Applications
- c. Case Study : SOKOA

## DAY 2 : Artificial Intelligence and Big Data

- a. What is Changing?
- b. The Underlying Technologies
- c. Lab : See and use a Big Data Platform

## DAY 3 : How to Successfully Run a AI/Big Data Project

- a. The 3 “Pillars”
- b. Project : “Virtual Salesman” Marketing Strategy and Sales Pitch

# DISCUSSION

*ARTIFICIAL INTELLIGENCE, BIG DATA, HAVE  
YOU EVER HEARD OF IT???*

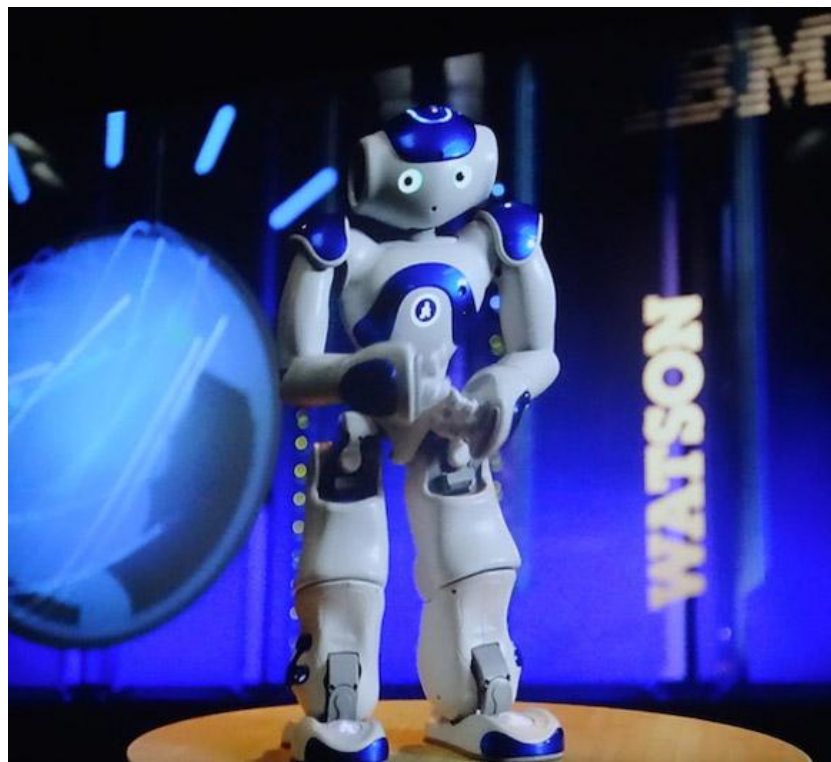


# YOU HEARD ABOUT DEEPLUE (1990)



*CHESS CHAMPION 1997*

# SAY HELLO TO WATSON (2017) !!!



<https://www.youtube.com/watch?v=L5QJs6byoaI>  
[https://www.youtube.com/watch?v=\\_Xcmh1LQB9I](https://www.youtube.com/watch?v=_Xcmh1LQB9I)

# AND WAYMO !!!



<https://www.youtube.com/watch?v=qtApzKnGU94>

# PROGRAM

## DAY 1 : Business Applications

- a. General Introduction : The New “Black Gold”
- b. Big Data for All !!!! - Examples of Real Applications
- c. Case Study : SOKOA

## DAY 2 : Artificial Intelligence and Big Data

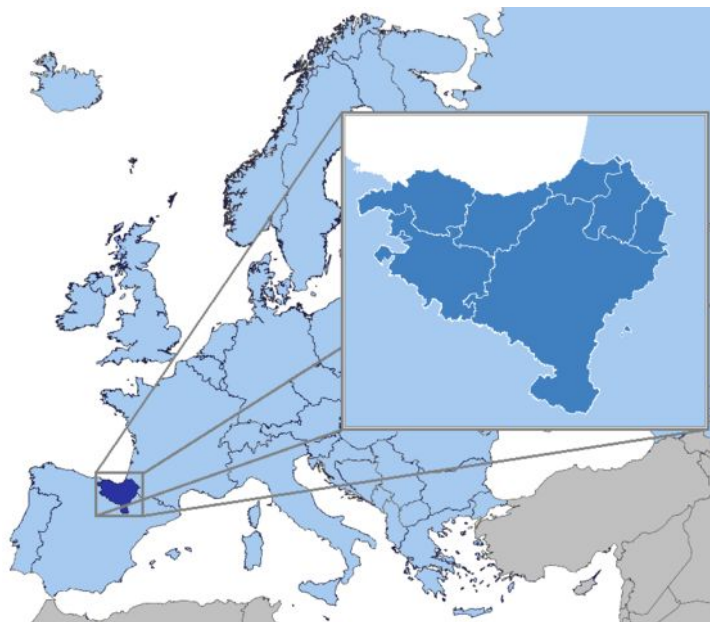
- a. What is Changing?
- b. The Underlying Technologies
- c. Lab : See and use a Big Data Platform

## DAY 3 : How to Successfully Run a AI/Big Data Project

- a. The 3 “Pillars”
- b. Project : “Virtual Salesman” Marketing Strategy and Sales Pitch



# HUPI PAYS BASQUE



# ABOUT HUPI

- Software company specialized in Artificial Intelligence and “Big Data”
- Created in 2014 (after 2 years of incubation)
- S.A.S with a capital de 373 840€
- “HUPI” platform designed to process large amounts of data in real time
- An Expertise in machine learning



Located on the Izarbel Technopole, Bidart



Tomorrow within the BIC of San Sebastian

**Euskadi - Laureate 2016**

**Concours “Industry 4.0”**



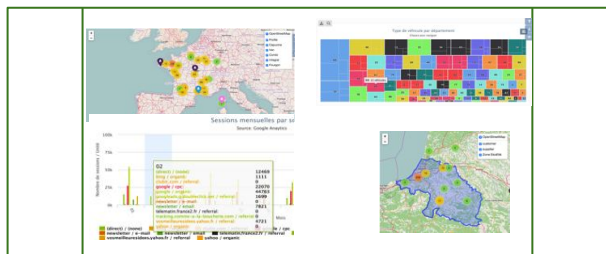
# FROM VISUALIZATION RECOMMENDATIONS

=> TO AUTOMATED

*Data Lake*



*Synthesis*



**Visualization**



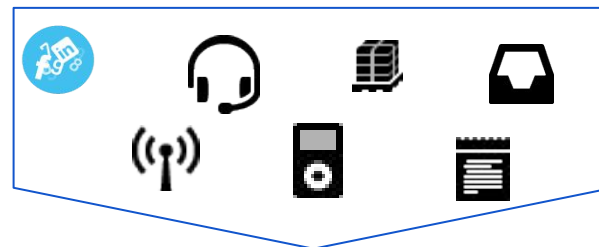
*Decisions*



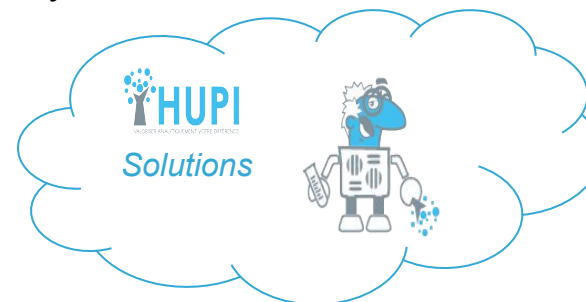
*Actions*



*Data River*



*Synthesis*



**Recommendations**



*Decisions*



*Actions*



# VALUE PROPOSITION

Creates and operates solutions, which generate recommendations in automatic and in real time, from you data, and integrate the results in your systems

These solutions use machine learning algorithms, developed by our own data scientists, and run on our proprietary Big Data platform, hosted in Bidart





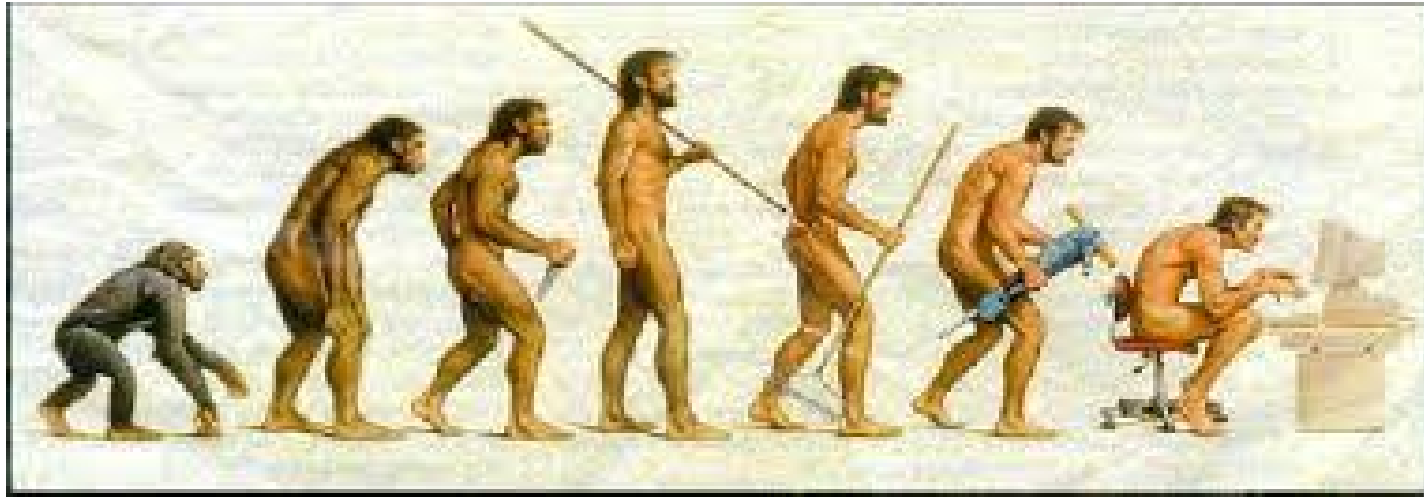
BIG DATA...

IS LIKE SEX FOR TEENAGERS...

- EVERYONE TALKS ABOUT IT...
- NO ONE KNOWS HOW TO DO IT...
- EVERYONE THINKS THAT EVERYONE DOES IT...
- SO EVERYONE SAYS THAT THEY ARE DOING IT...

# THE BIG DATA REVOLUTION...

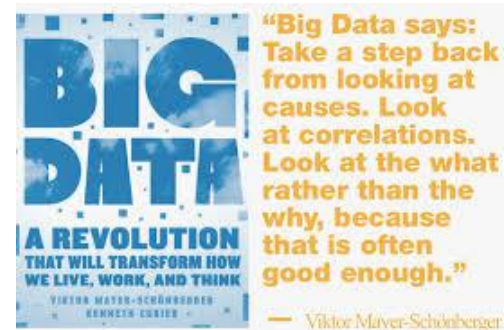
... IS AN EVOLUTION



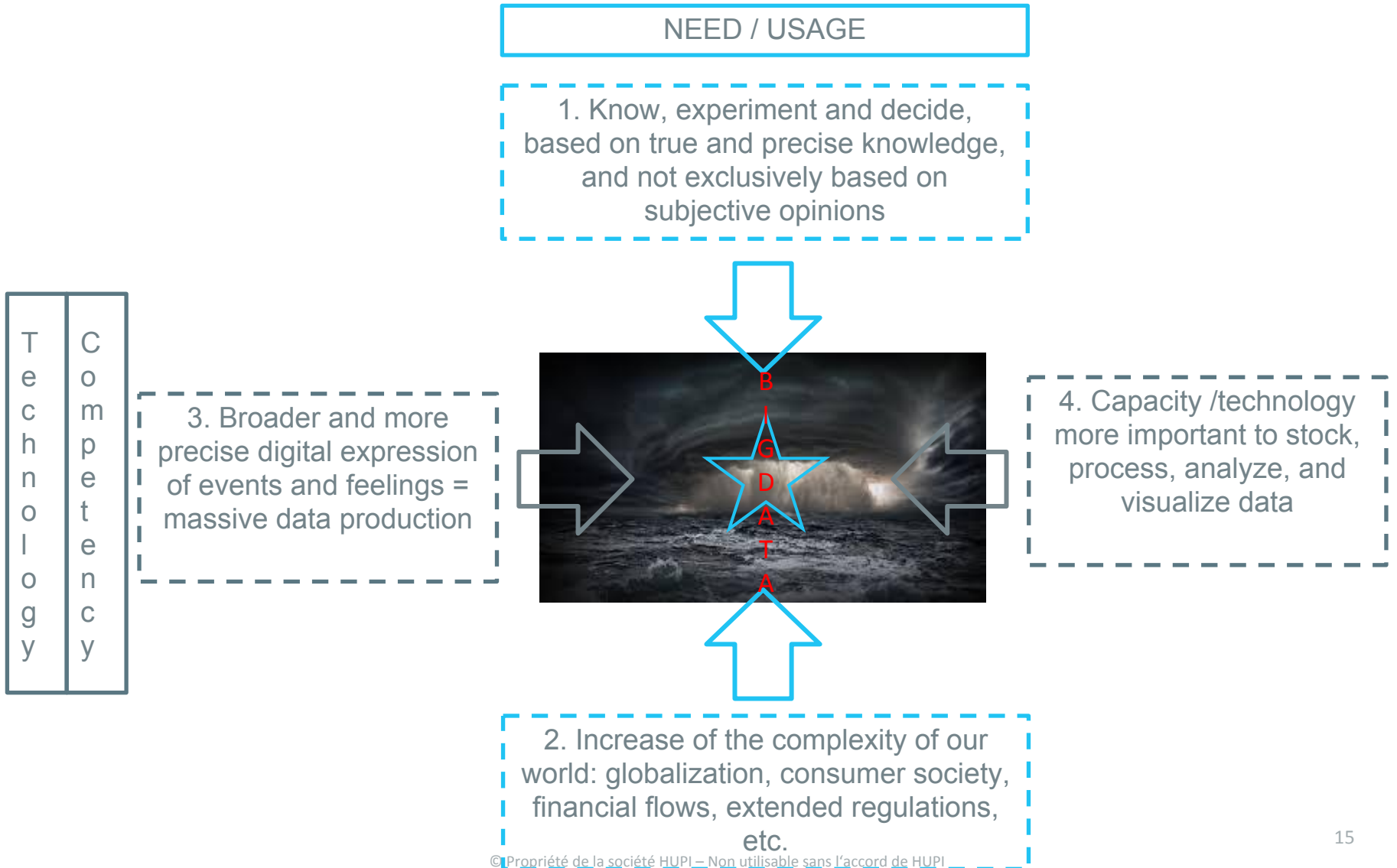
**Damn!**

**We have to do « with it » because « it is going to last »**

- As important as the arrival of the Internet
- An opportunity to reshuffle the cards



# THE PERFECT STORM : A DISRUPTIVE PHENOMENON EMERGING FROM 4 FORCES



# PLACE THE DATA AT THE CENTER OF THE DECISION MAKING PROCESS



## 1. Transparency

Allow an understanding of the internal and external events that is more accurate, more precise, more complete, more comprehensive and above all more contextualised



## 2. Decision

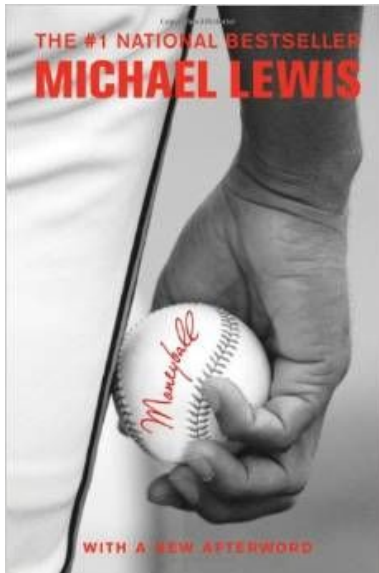
Help or automate some decision making processes



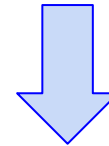
## 3. New services

Experiment and implement new innovative services that have their own economical models

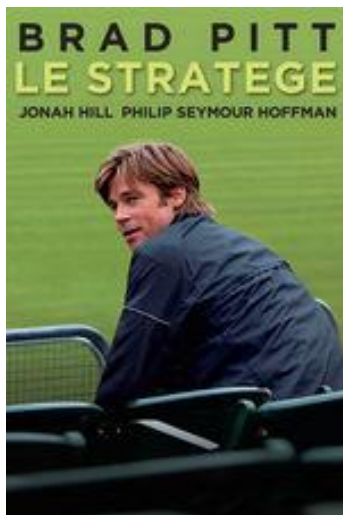
# MONEYBALL : THE ART OF WINNING AN UNFAIR GAME



The collective knowledge of baseball professionals from the last century is subjective and fallible, and do not allow to reflect the real level/value of the players.



Manage sport teams through the mathematical analysis of their behaviour rather than exclusively through experience and intuition

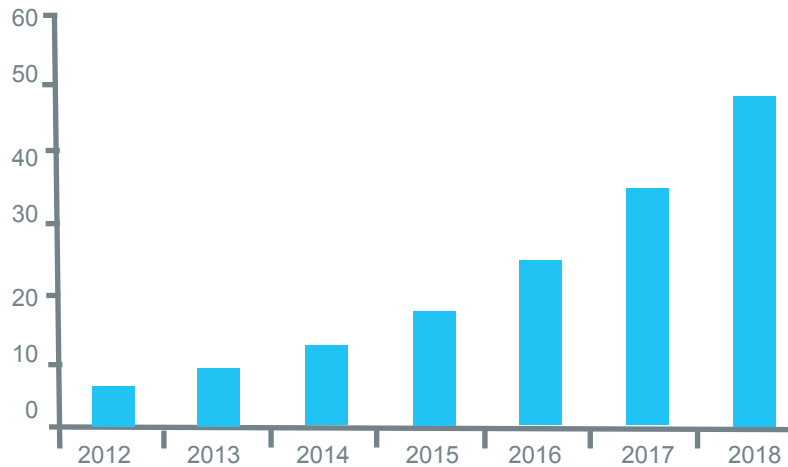


- <https://www.youtube.com/watch?v=AiAHIZVgXjk>
- <https://www.youtube.com/watch?v=TpBcwGOvO80>
- <https://www.youtube.com/watch?v=yGf6LNWY9AI&list=RD>  
[TpBcwGOvO80&index=2](https://www.youtube.com/watch?v=TpBcwGOvO80&index=2)

## CORE BELIEF

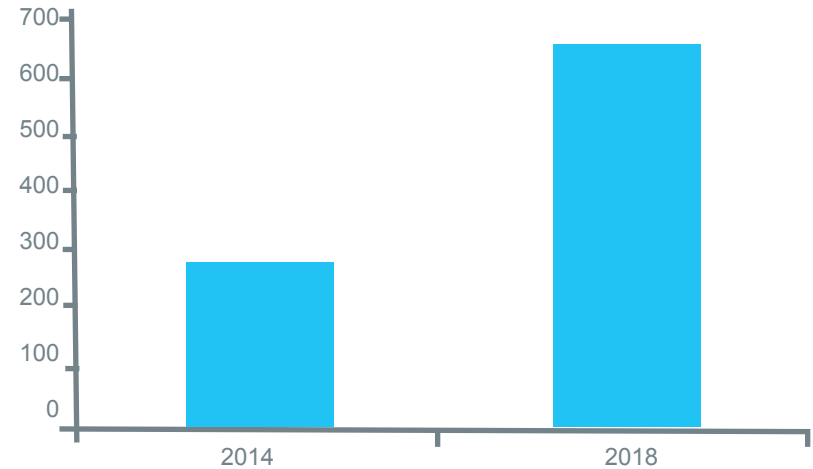
1. Big Data is not about technology, but about the real-time use of data in the front-line execution – ***Data without execution does not bring any value***
2. The most important success factor is the man-machine interface and the decisions made by humans based on analytics insight
3. Any Big Data solution has to be « business first », hypothesis driven, rather than trying to gather all the data available  
– ***Half the battle is to identify and eliminate the data that is irrelevant***
4. Now matter how incomplete your data is right now, you can nevertheless use it better to create business value through the use of Big Data Tools  
– ***no « Apollo » project is needed to get going***

# A MAJOR STEP FORWARD FOR ALL ENTERPRISES...



*Global Market (in \$ Billions)*

A global market valued at \$48,3 Bn in 2018



*French Market (in \$ Millions)*

A growth of the French market estimated at +129% between 2014 and 2018

# ... EVERY SECTOR CAN BENEFIT

## Les initiatives Big Data en forte progression

### Adoption et Projets Big Data

Evaluation et  
Projet Big Data












**20%**

**56%**  
Pas d'initiatives  
Big Data

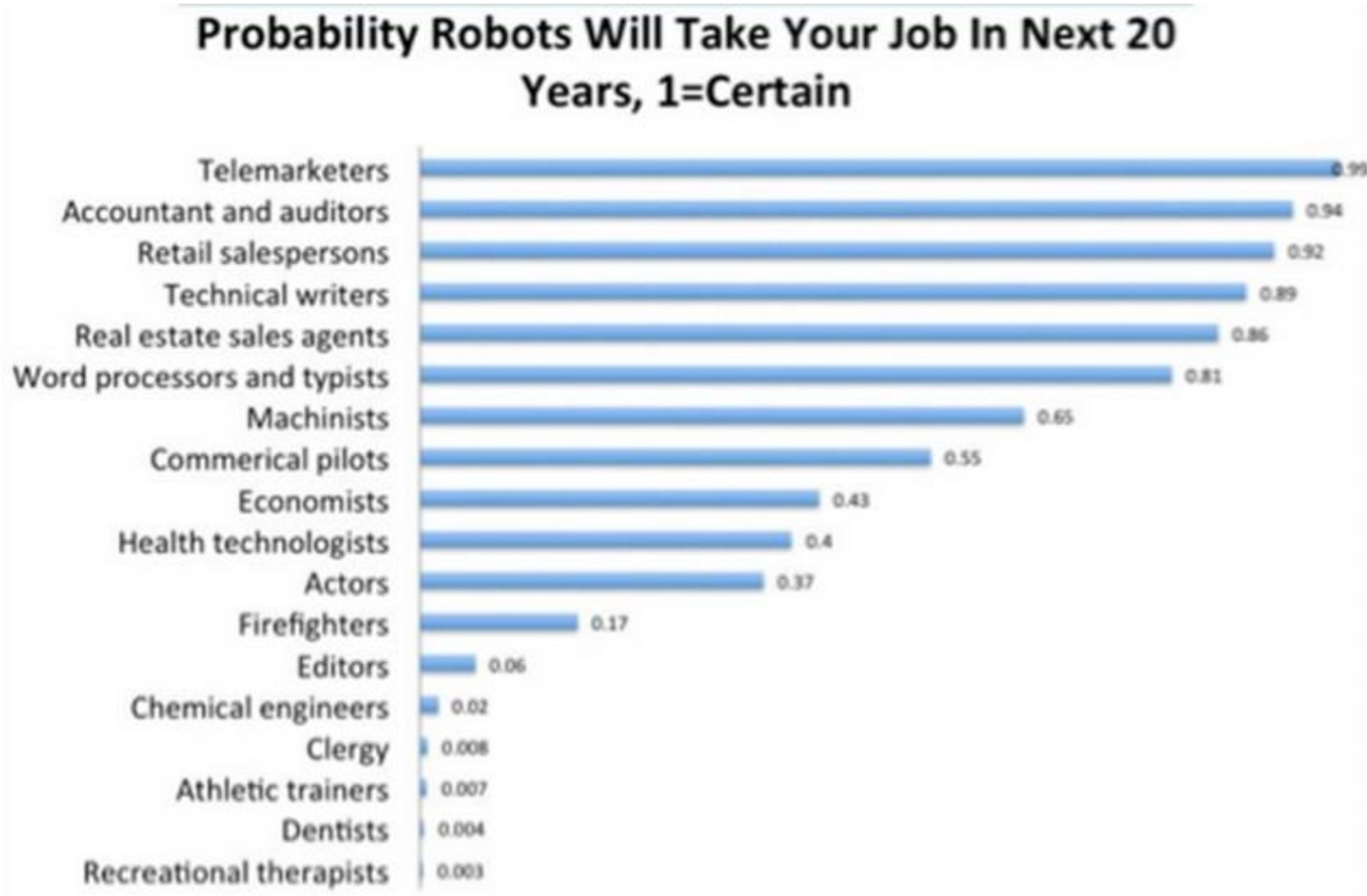
**24%**

Adoption  
Big Data  
(solution  
en place /  
en cours)

### Adoption par secteurs d'activité

	Commerce	<b>33%</b>
	Transports	<b>30%</b>
	Services Financiers	<b>30%</b>
	Distribution	<b>29%</b>
	Santé	<b>25%</b>
	Manufacturing	<b>24%</b>
	Services	<b>24%</b>
	Utilities	<b>22%</b>
	Télécoms	<b>22%</b>
	Secteur Public	<b>17%</b>
	Education	<b>16%</b>

# ... JOBS WILL BE IMPACTED



# EXAMPLE OF THE CNCC



# Solution de Veille Automatique Intelligente pour la Gestion des Risques



**Objectif:** Disposer d'un système de **Veille Automatique et Intelligent** permettant, pour chaque société de son portefeuille, de :

- remonter les informations externes pouvant avoir un impact sur l'activité du CAC
- croiser et analyser des informations issues de plusieurs sources : informations légales mais aussi différents réseaux
- détecter automatiquement des points d'attention devant être portés à l'attention du Commissaire aux Comptes dans le cadre de sa mission
- d'alerter sur des événements nouveaux, « suspects » ou « inhabituels »
- de scorer par importance/pertinence/fiabilité les éléments mis à disposition du CAC

# Bénéfices de la Solution de Veille pour les CAC



Accès à des Informations Dissimulées

Veille active et systématique via une recherche “continue” d’informations dispersées aux 4 coins du web et parfois dissimulées (journaux, sites, communiqués, etc.)

Vue “croisée” de l’entreprise

Remontée et croisement d’informations, issues de nombreuses sources de données, de types variées, pour une vision “croisée” et rechercher les incohérences (clients, fournisseurs, etc.)

★ Système d’Alertes Temps réel

Génération d’alertes, ultra qualifiées (priorisées par niveau de sévérité), sur des risques peu apparents, liés à l’entreprise

★ Qualification fine des informations remontées

Organisation et priorisation fine, au sein d’une application dédiée, de l’ensemble des informations remontées, afin de mettre en relief ce qui semble pertinent et important

Historisation des informations

Accès à l’historique des informations remontées pour des recherches a posteriori

Automatisation de l’ensemble du processus

Toutes ces informations et alertes sont poussées automatiquement, en continu, qualifiées et pré-organisées, de manière ultra-personnalisée, à chaque CAC, sans intervention de celui-ci

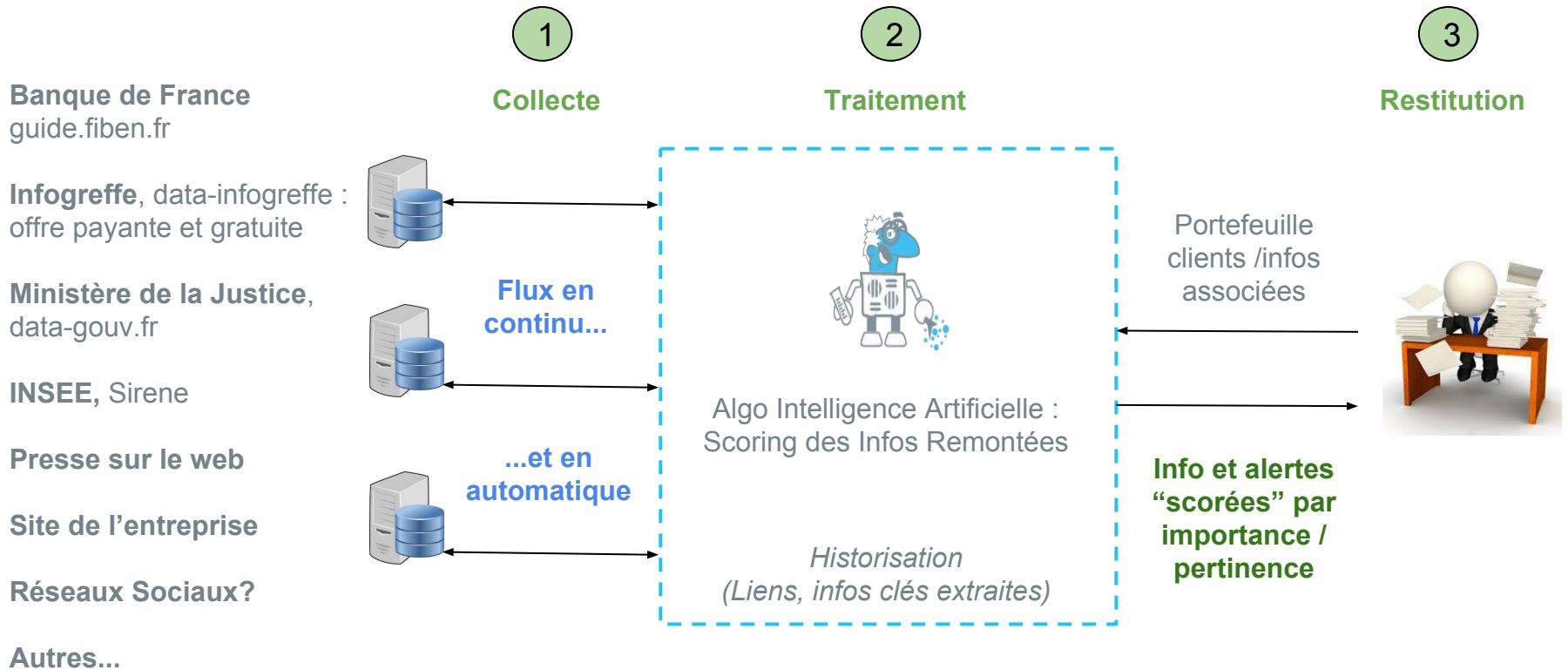
# Schéma Fonctionnel



## Sources de Données

## Plateforme HUPI

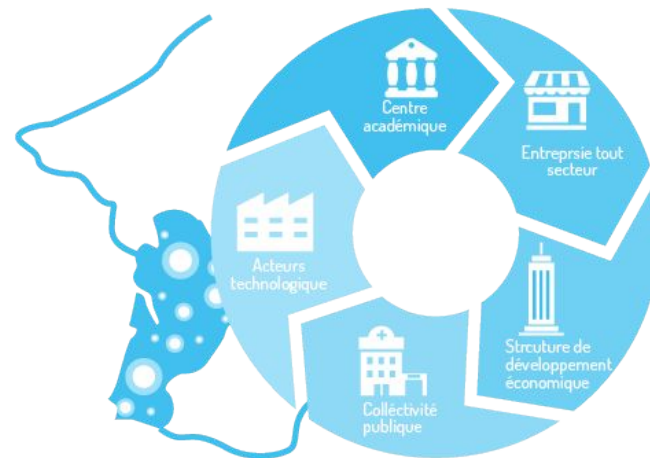
## Application pour les CAC



QUESTIONS ?



## b. BIG DATA FOR ALL



# APPLICATIONS EVERYWHERE

- Predictive maintenance or condition monitoring
- Warranty reserve estimation
- Propensity to buy
- Demand forecasting
- Process optimization
- Telematics

## Manufacturing



- Predictive inventory planning
- Recommendation engines
- Upsell and cross-channel marketing
- Market segmentation and targeting
- Customer ROI and lifetime value

## Retail



- Alerts and diagnostics from real-time patient data
- Disease identification and risk stratification
- Patient triage optimization
- Proactive health management
- Healthcare provider sentiment analysis

## Healthcare and Life Sciences



- Aircraft scheduling
- Dynamic pricing
- Social media – consumer feedback and interaction analysis
- Customer complaint resolution
- Traffic patterns and congestion management

## Travel and Hospitality



- Risk analytics and regulation
- Customer Segmentation
- Cross-selling and up-selling
- Sales and marketing campaign management
- Credit worthiness evaluation

## Financial Services



- Power usage analytics
- Seismic data processing
- Carbon emissions and trading
- Customer-specific pricing
- Smart grid management
- Energy demand and supply optimization

## Energy, Feedstock, and Utilities



## ...REALLY EVERYWHERE...

- SCIENTIFIC RESEARCH (MEDICAL, ENVIRONMENTAL, ...)
- EDUCATION
- GOVERNMENT POLICIES
- GOUVERNANCE
- SMART CITY
- ETC.

# ...EVEN IN THE SEARCH FOR LOVE



Match.com has a group of data scientists improving continuously the 15 algorithms « matching » entre people





# FOCUS ON E-COMMERCE

*Our Partners :*



# FOCUS ON E-COMMERCE: VISITS, VISITS BUT FEW PURCHASES

98% of the visitors of an e-commerce website do not make a purchase

85% of the visitors who do not find their product in less than 2 minutes leave the site...



... and 60% of them will never come back

# WELCOME TO HUPI SALESMAN

Optimize the conversion rate and the average basket size...  
... through the « Ultra-personalization » of the client relationship in real time...  
... by showing the right product, at the right time, to the right visitor

30% of  
Amazon sales



With the Virtual Salesman



Personalized product recommendation,  
in automatic and in real time



# WIKICAMPERS



## MÉTIER

e-commerce website specialized in peer to peer rental of camping-cars

## BESOINS

- Increase the portfolio of vehicles available for rent by providing new services to the owners
- Increase the conversion rate, the probability to rent a camping car at each visit

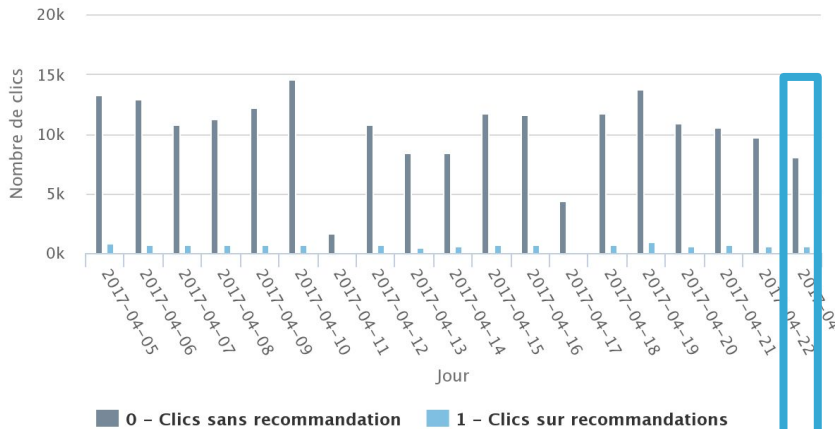
# PERFORMANCES



7.0  
% Reco. validées



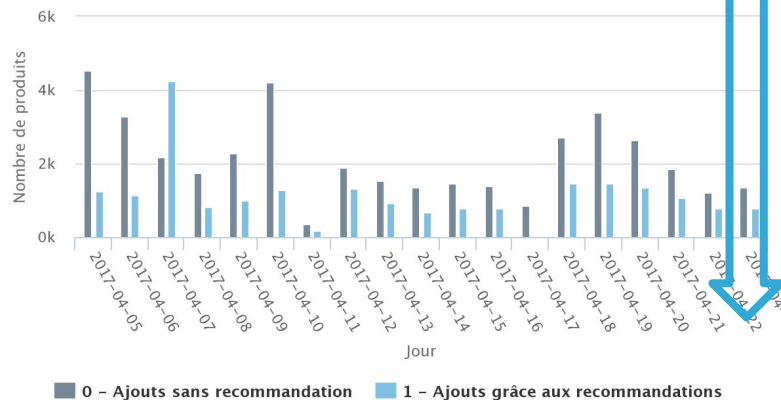
Nombre de clics journaliers sur les produits non recommandés et Nombre de clics journaliers sur les produits recommandés



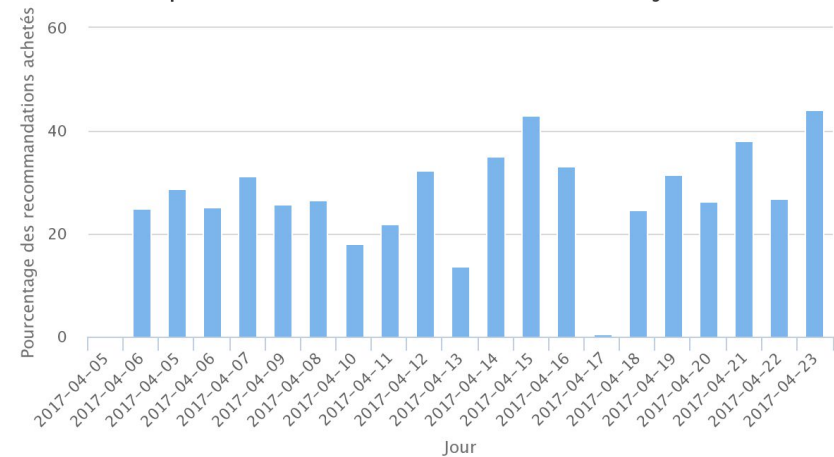
## Tunnel

- 8 600 click on products from which “only” 590 clicks coming from HUPI recommendations
- FOR 2154 products purchased from which 779 products coming from HUPI recommendations

Nombre d'ajouts au panier journaliers sans recommandation et Nombre d'ajouts au panier journaliers grâce aux recommandations



Part des produits recommandés dans les achats journaliers



Series 1

## PROJECT : VIRTUAL SALESMAN (1/3)

### Objective:

- develop your recommendations to design an efficient Marketing Strategy to scale-up the sales of “HUPI Virtual Salesman” product
- prepare a 5mn pitch, with responses to objections, to sell HUPI Virtual Salesman to the e-commerce firm Scandivie.



### Presentations in groups of 2-4 : Friday October 27th afternoon

- Marketing Strategy: 30mn
- Sales Pitch: 5mn

## PROJECT : VIRTUAL SALESMAN (2/3)

### Key Attributes of HUPI VIRTUAL SALESMAN :

- Recommendations displayed on the Homepage, category page, product page, basket page; several underlying algorithms all using machine learning techniques
- No setup fee, just monthly fees with no commitment : 100€ - 400€/month depending on e-commerce actual turnover
- Technical integration: light and rapid (< 2 days); service operational within 1 week
- Monthly reports provided on actual performance

# PROJECT : VIRTUAL SALESMAN (3/3)



<https://www.scandi-vie.com/>

## BUSINESS

Pure e-commerce player,  
created in 2000,  
distributing high-end kitchen utensils,  
made in Scandinavia.

Large portfolio of products proposed

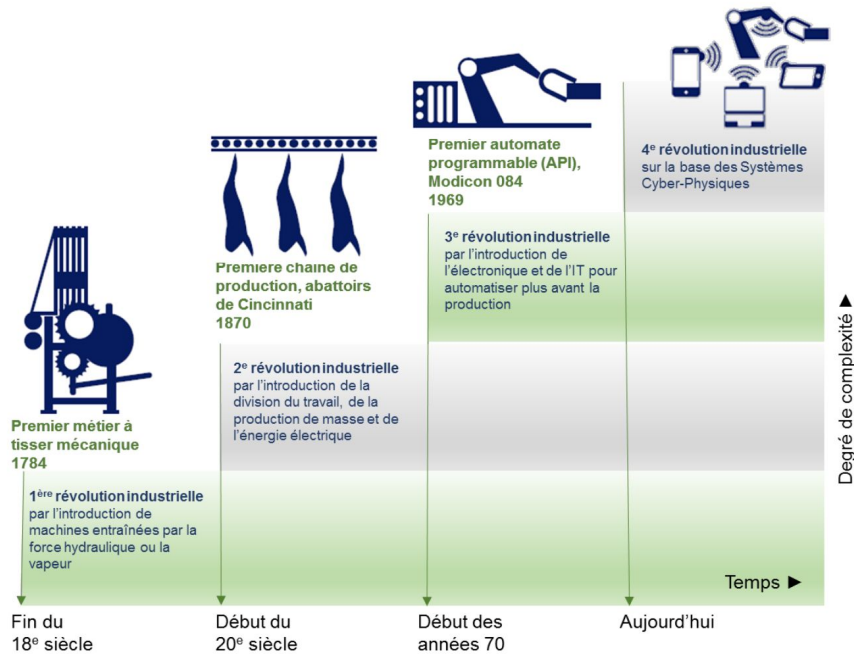
More than 100+ visitors each day



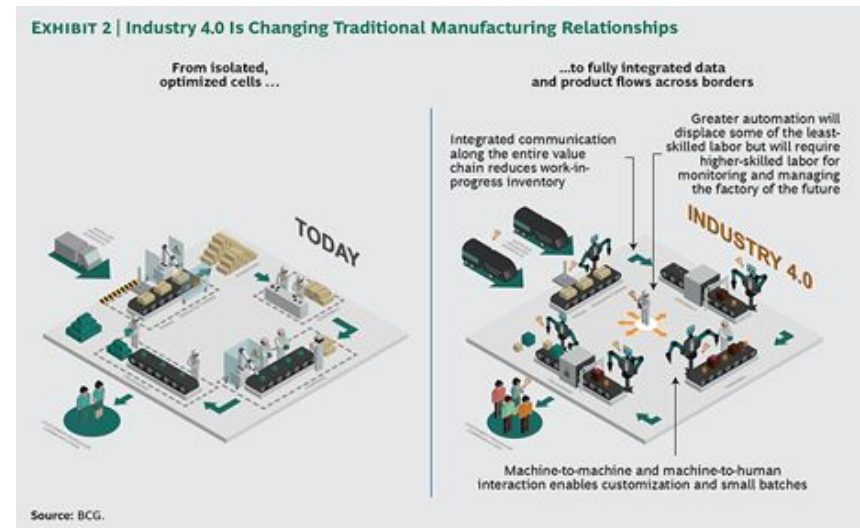
# FOCUS ON INDUSTRY 4.0

# INDUSTRY 4.0 = 4th Industrial Revolution

Figure 1 : La quatrième révolution industrielle



Source : ©DFKI, 2011



# THE EXAMPLE OF SIEMENS (Video)

**SIEMENS**

<https://www.youtube.com/watch?v=HPRURtORnis&t=26s>

## BIND 4.0: THE EXAMPLE OF EUSKADI



[https://www.youtube.com/watch?time\\_continue=1&v=YKQpKYMsTY](https://www.youtube.com/watch?time_continue=1&v=YKQpKYMsTY)

W

# 10 CAS D'USAGES DE L'INDUSTRIE 4.0

**EXHIBIT 1 | Ten Use Cases Show the Effects of Industry 4.0 on the Workforce**

	<b>BIG-DATA-DRIVEN QUALITY CONTROL</b> Algorithms based on historical data identify quality issues and reduce product failures
	<b>ROBOT-ASSISTED PRODUCTION</b> Flexible, humanoid robots perform other operations such as assembly and packaging
	<b>SELF-DRIVING LOGISTICS VEHICLES</b> Fully automated transportation systems navigate intelligently within the factory
	<b>PRODUCTION LINE SIMULATION</b> Novel software enables assembly line simulation and optimization
	<b>SMART SUPPLY NETWORK</b> Monitoring of an entire supply network allows for better supply decisions
	<b>PREDICTIVE MAINTENANCE</b> Remote monitoring of equipment permits repair prior to breakdown
	<b>MACHINES AS A SERVICE</b> Manufacturers sell a service, including maintenance, rather than a machine
	<b>SELF-ORGANIZING PRODUCTION</b> Automatically coordinated machines optimize their utilization and output
	<b>ADDITIVE MANUFACTURING OF COMPLEX PARTS</b> 3-D printers create complex parts in one step, making assembly redundant
	<b>AUGMENTED WORK, MAINTENANCE, AND SERVICE</b> Fourth dimension facilitates operating guidance, remote assistance, and documentation

Sources: Expert interviews; BCG analysis.

Des algorithmes pour améliorer la qualité au travers de l'analyse fine des données

Des robots humanoïdes facilement formés et entraînés pour réaliser de nouvelles tâches

Un système de transport de composants dans l'usine indépendant et intelligent

Des logiciels de simulation de lignes de production avant installation

Des logiciels pour mieux surveiller l'ensemble de la chaîne logistique

Des systèmes pour alerter un fabricant de la nécessité d'opérer une action de maintenance avant toute casse

Vendre des services au lieu de vendre le composant réalisant le service

Des systèmes intelligents qui se coordonnent automatiquement afin de maximiser l'utilisation de chaque machine

Les imprimantes 3D ou les lasers intelligents permettent de créer des composants complexes en une étape

Des systèmes pour mieux localiser des composants et permettre une assistance à distance

## Exemples



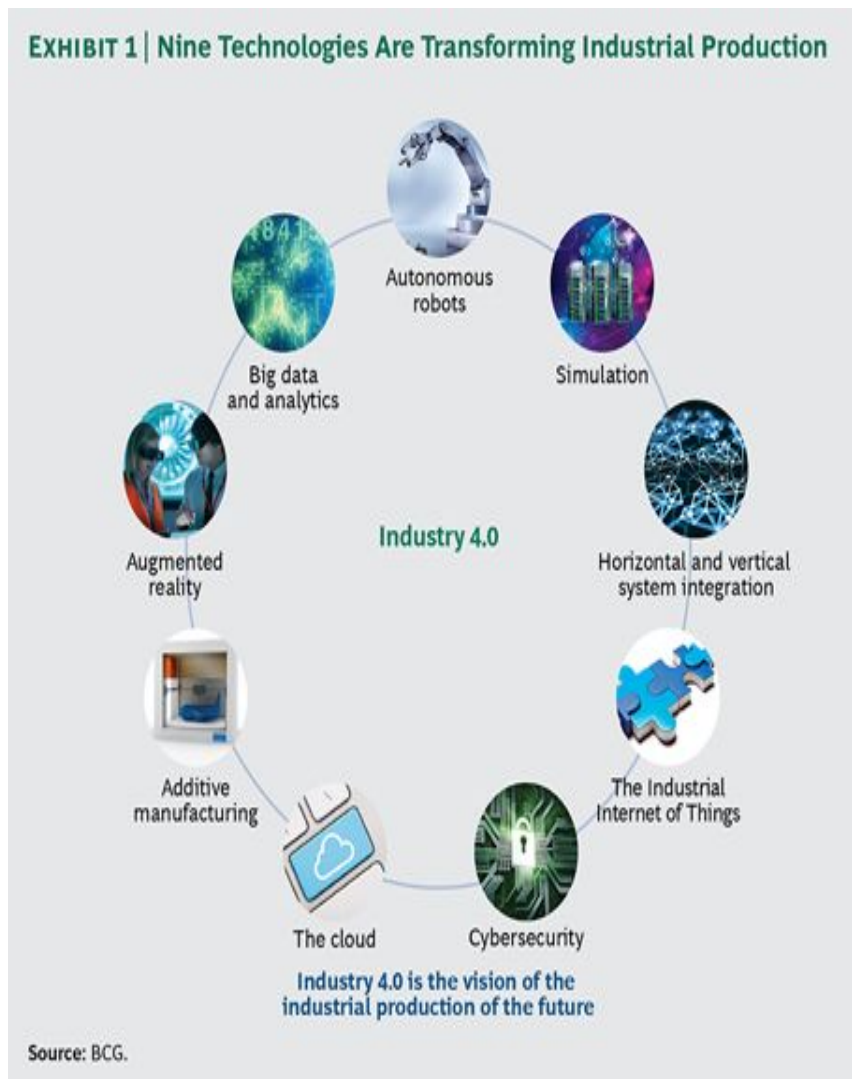
Mercedes-Benz  
Vito - Sprinter



# 9 TECHNOLOGIES FOR THE INDUSTRY 4.0



EXHIBIT 1 | Nine Technologies Are Transforming Industrial Production



**All HUPI Partners**

**All HUPI Partners**



## YUMI (You and Me): COBOT



### The 1st robot who works along men without a protecting cage

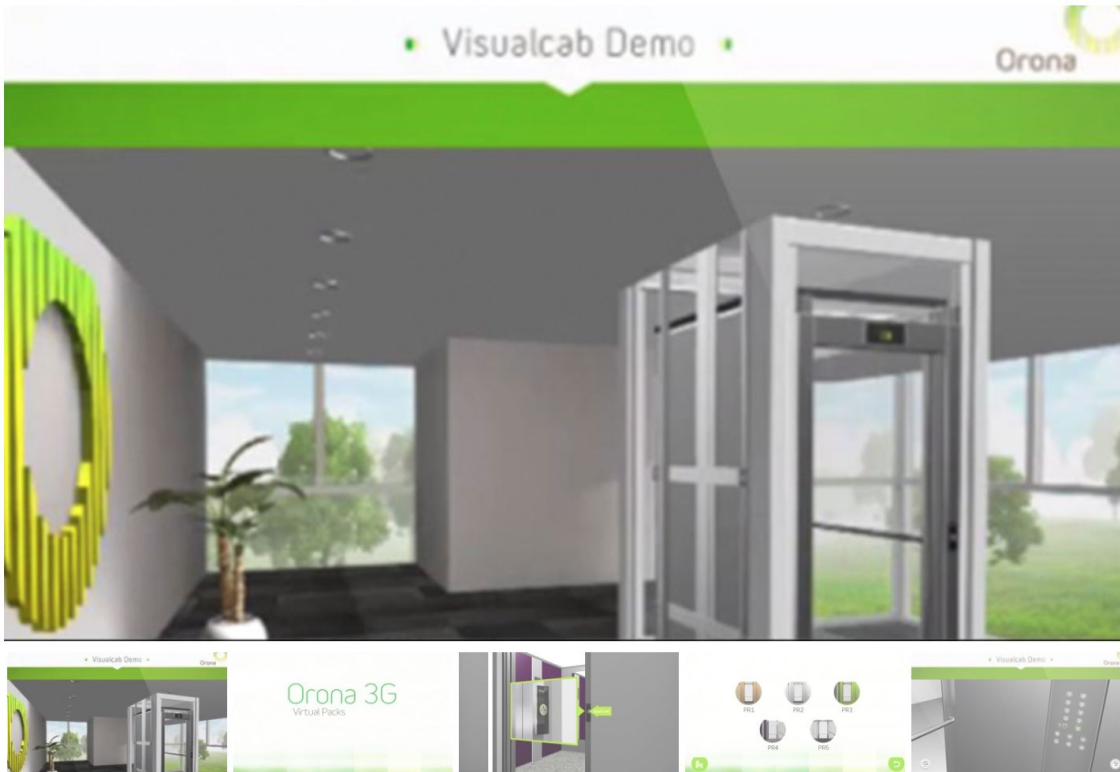
- More autonomous, more flexible, and more cooperative
- Designed to interact with other robots and learn from humans
- Designed to assemble products along human
- Manipulates with very high precision any kind of small objects
- With 2 physical arms
- A vision to ensure a safe interaction and recognize the components to assemble

<http://new.abb.com/products/robotics/fr/robots-industriels/yumi>



Augmented reality

# AUGMENTED REALITY



A tool to enable a sales team to configure live an elevator in front of a client.

The application creates 3D images which allow the client to have a immersed view of the elevator chosen

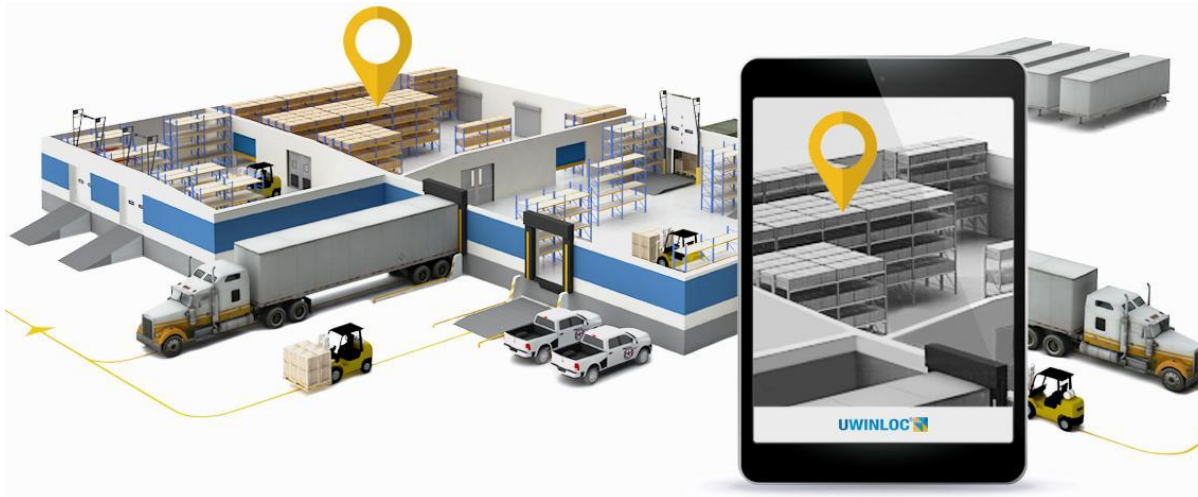
<http://www.innovae.eu/projects/soraluce-imts/>





## To locate with precision large quantities of objects

The UWINLOC solution enables to locate in 3D and identify with a precision of 30 cm any kind of objects, inside or close to a building, using battery-less stickers, worth a few cents of euros





# SIMULATION



## Modeling the pilot's consciousness of the situation



Application of a simulator to model the actual activity of a plane pilot and design interfaces presenting a content adapted to his capacity to manage the situation



# SOME CLIENTS USE CASES



# ENEDIS: The Virtual Assistant of the Mobile Technician

**P5**

Recharger

OUVRIRE IPS 4.1

**Point d'observations**

Chaque cartouche fusible doit être fixée sur un protecteur porte-fusible.

**GTDE**

Fiche 2.1.2 et 3.3.12

**Opérations - Risques**

Insertion excessive de cartouche, fusible ou d'accessoire - risque de déplacement de circlips (échauffement ou court-circuit)

**Mesures à prendre**

Utilisation obligatoire du protecteur porte fusible pour les cartouches fusibles. Util d'outillage limitant la pénétration des couteaux dans les contacts.

**Cas d'erreurs**

VALIDER REFUSER

**HUPI**

Recherche Hupi

- Ce PDC comporte 3 risques identifiés -IDC16410163-ANGLET 2017-05-19  
Ce PDC comporte 3 risques identifiés Risques routiers Risques d acces Autres dangers....
- Ce PDC comporte 2 risques identifiés -IDC16402895-HENDAYE 2017-05-18  
Ce PDC comporte 2 risques identifiés Risques routiers Risques d acces....
- Ce PDC comporte 5 risques identifiés -IDC16410166-SEIGNOSSE 2017-04-25  
Ce PDC comporte 5 risques identifiés Risques routiers Risques d acces Risques Environnement Co activite Produits dangereux....
- Ce PDC comporte 2 risques identifiés -IDC16404130-SOUSTONS 2017-04-10  
Ce PDC comporte 2 risques identifiés Risques d acces Geste et posture....
- matthieu laffaille 2017-05-19  
matthieu laffaille enedis grdf fr....
- matthieu laffaille 2017-05-18  
matthieu laffaille enedis grdf fr....
- nicolas a bonnafous 2017-05-16  
nicolas a bonnafous enedis grdf fr....
- Michael Robin 2017-04-28  
michael robin enedis grdf fr....
- Matthieu m Laborde 2017-04-28



Kit Main libre / Google Glass

Caméra thermique + Appareil photo

Bracelet tablette + GPS + Vibreur

Capteur de distance + Capteur de mouvement

**ENEDIS**  
L'ELECTRICITE EN RESEAU

# SAFRAN: Ultra personalize maintenance contracts for Helicopter Engines leveraging environmental data

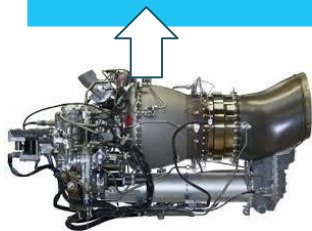


Données d'usages  
**externes**/internes opérateurs

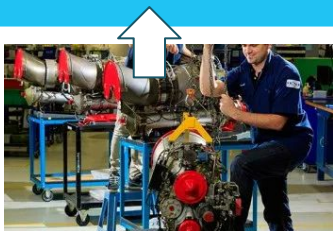


Données **externes**  
environnementales (météo,  
topologie, pollution, ....)

COLLECT / MODELING



Données internes moteurs

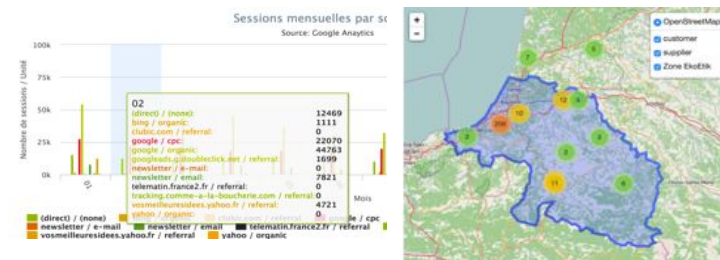


Données internes maintenance

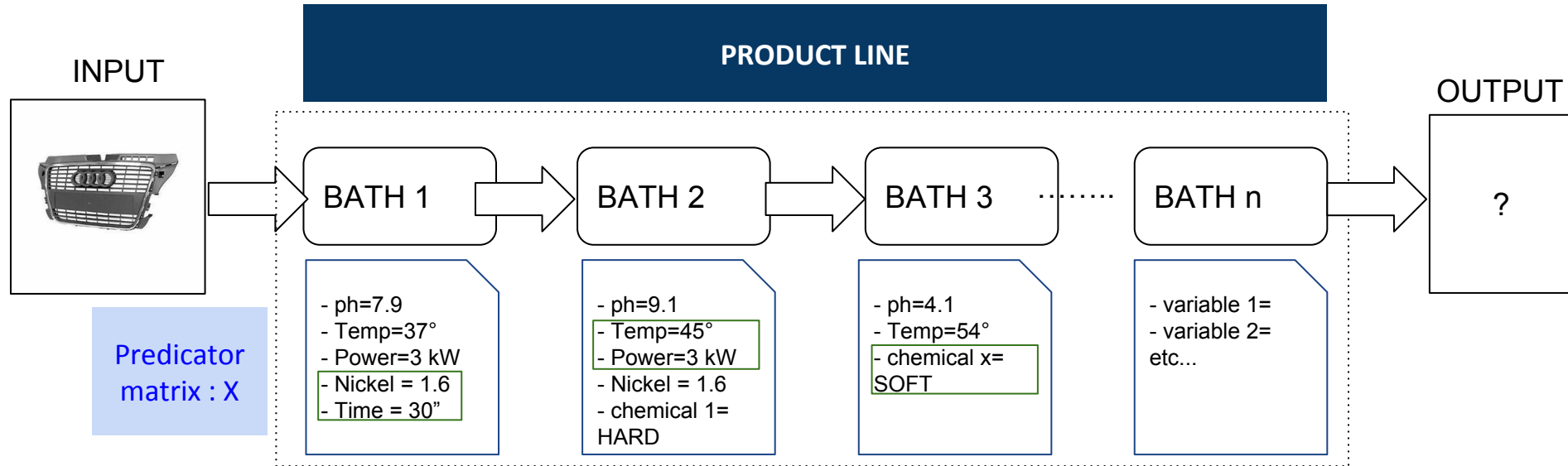


Coûts de maintenance / Type  
de contrat interne

Application  
(possible)



# MAIER: Optimization of a Manufacturing Chain



With these parameters on my product Line, what will be the quality of the products?  
*Identify main variables and predict quality impact*

Results matrix : Y



# AGUR: Nouveaux Services autour de la Gestion de l'Eau

- Optimisation des coûts de distribution
- Gestion de la ressource
- Qualité du services
- Alertes
- Prédictions

IA + IOT



AGUR fait partie du Groupe ETCHART (Groupe familial basque) de 1200 personnes spécialisé dans les métiers du BTP, de l'eau et de l'environnement.

# KEOLIS: Prédire la maintenance des bus à partir d'objets connectés embarqués

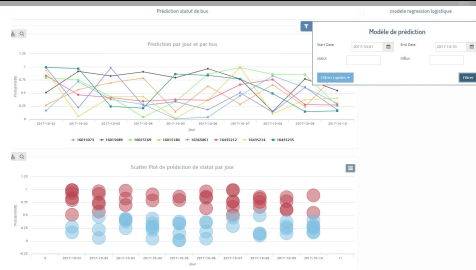
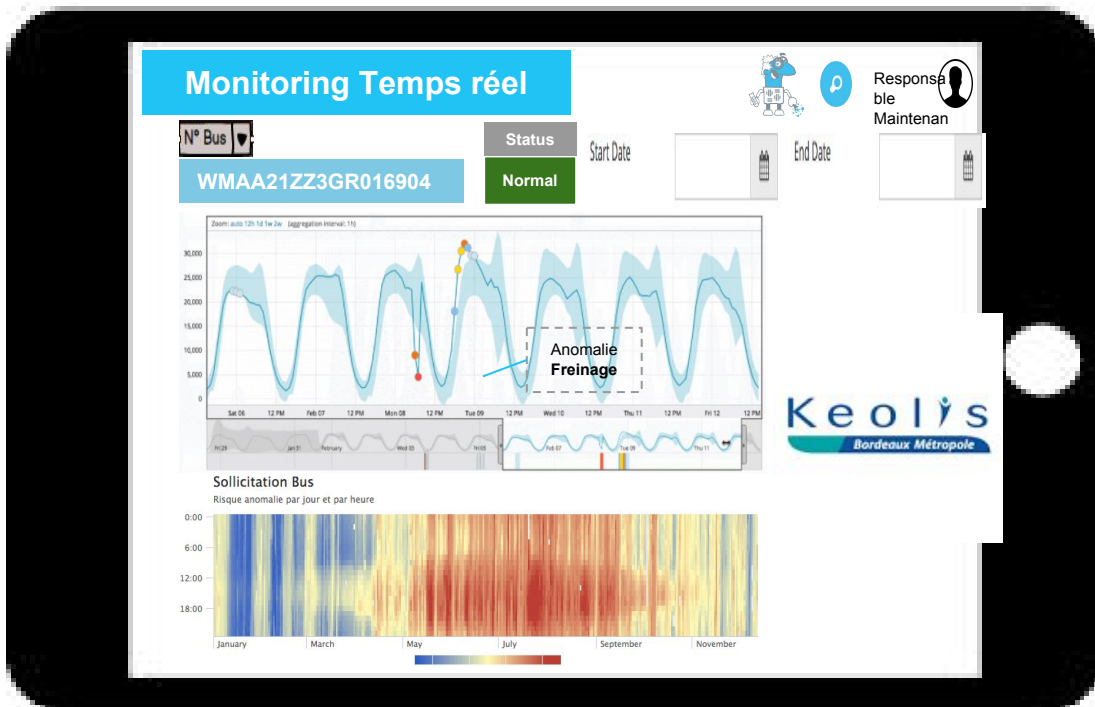
## Objectif:

-Anticiper les opérations de maintenance à réaliser sur chaque bus, en détectant, en continu, les “premiers signaux faibles” indiquant, avec un certain niveau de probabilité, un risque de panne à court terme

## IOT + IA

## Approche:

- Adopter une approche “statistique et probabiliste” (non “déterministe”), basée sur des données historiques précises et caractérisées
- Construire et opérer, en continu, un modèle auto-apprenant liant les données dites “comportementales” des bus et les défaillances effectivement constatées
- Développer une interface générant des alertes sur les bus évalués à “fort risque”

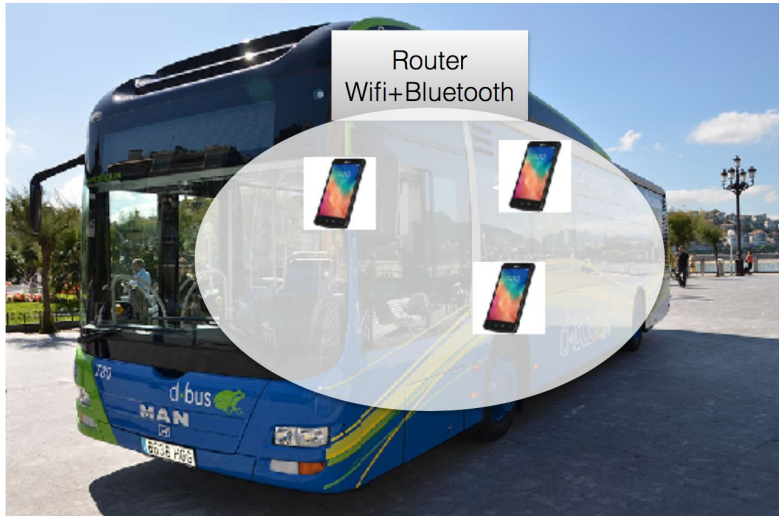


# LUMA : New services based on the geo-localization of bikes

Objective : Provide new services to bike riders, cities and equipment providers, based on data obtained from the geo-localization of bikes



# DBUS : Passengers Feedback in real time



## Objectif :

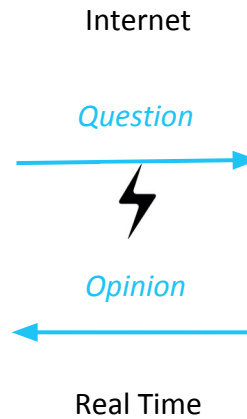
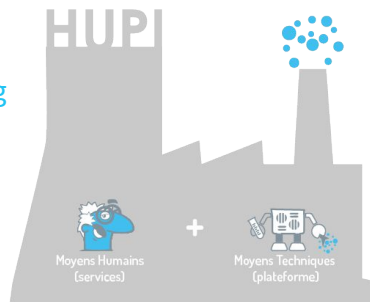
Allow passengers to simply provide in real time their opinion on a given topic related to DBUS services, that can be updated in automatic and at a given frequency



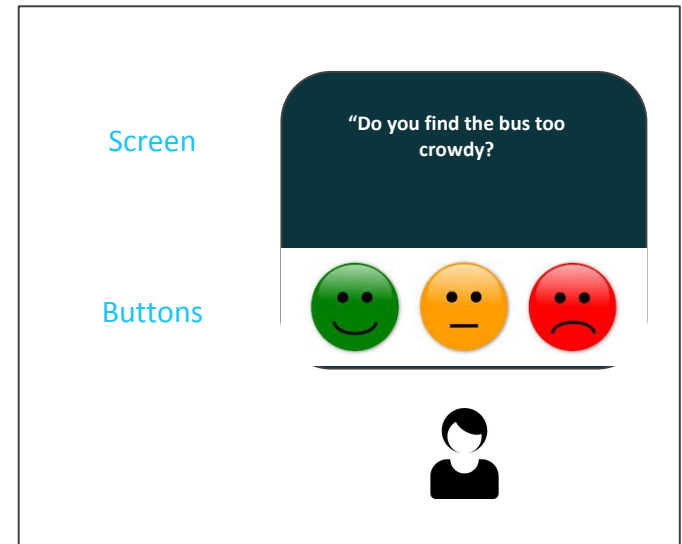
# HOW IT WORKS...

## HUPI ARTIFICIAL INTELLIGENCE ENGINE

Machine Learning Algorithm



## DEVICE EMBARKED IN DBUS



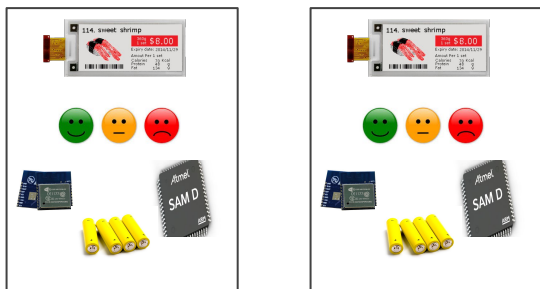
Data Sources

- Library of questions related to DBUS service quality
- Historic of opinions provided by passengers
- Geolocation of the bus
- Bus travel itinerary
- Weather
- Date



# TECHNICAL ARCHITECTURE

## DEVICES EMBARKED IN DBUS



3G-Wifi router embarked in Dbus



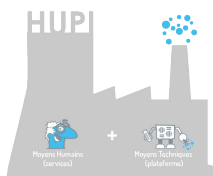
*Wakes up & connect to the Internet on every button click.*

*Connects to the Internet periodically.*

Internet

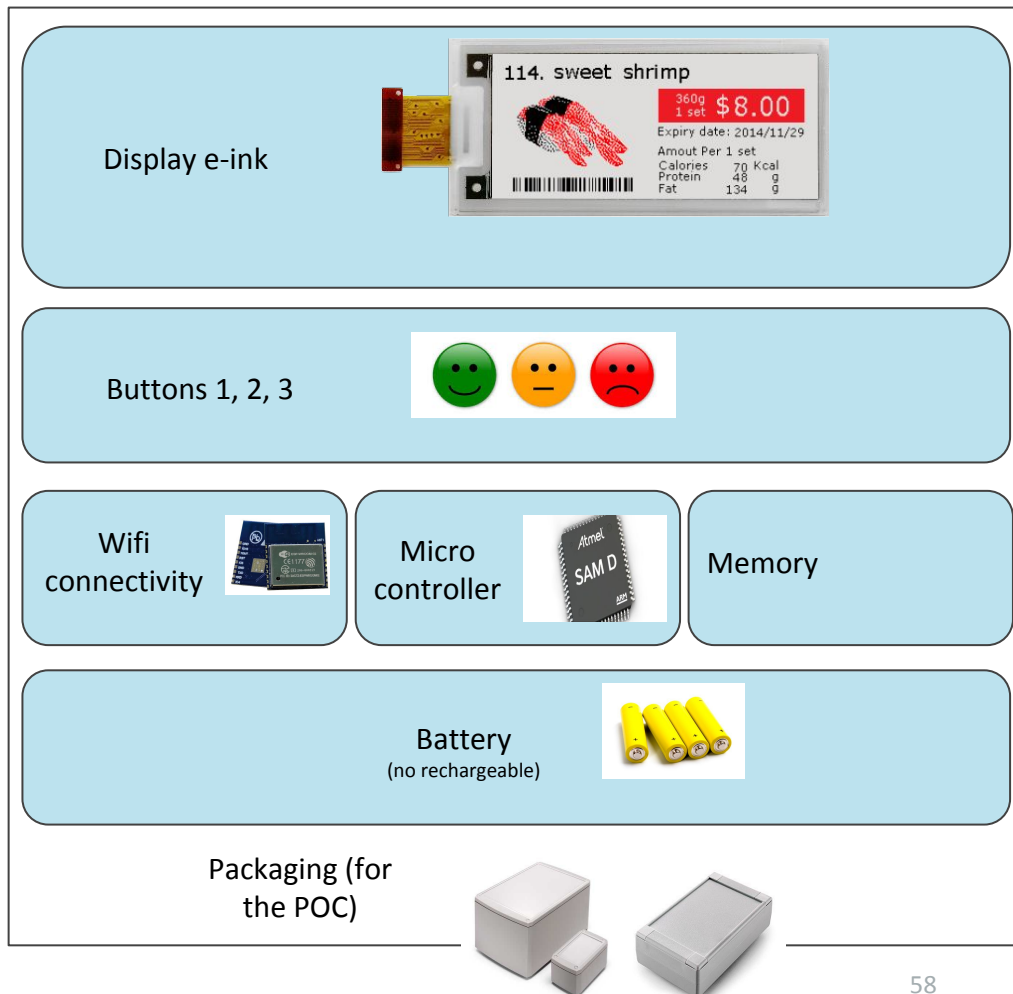


**HUPI ARTIFICIAL INTELLIGENCE ENGINE**



*Algorithms runs each time, HUPI receives an event*

## DEVICE EMBARKED IN DBUS



# OBJECTIVE: Optimize Fleet Management

Predict the amount of passengers expected for each line,  
at different periods of a given coming day

## Innovations :



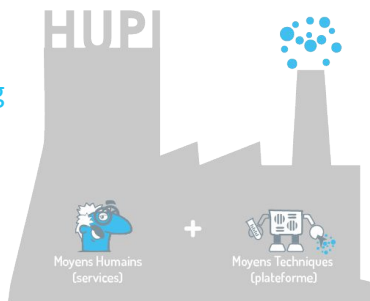
- **Leverage DBUS Controllers estimates :**  
 the machine learning algorithm will include data leveraging the experience of DBUS internal agents
- **Enrich training database with OPEN DATA :** additiona(weather, traffic,  
<http://www.donostia.eus/ataria/es/web/info/planak>)

Know how many passengers will probably board on a bus; at a given station, at a given moment of the following day, knowing...  
 past history, period of year, estimates provided by DBUS controllers, as well as the expected meteo for the day

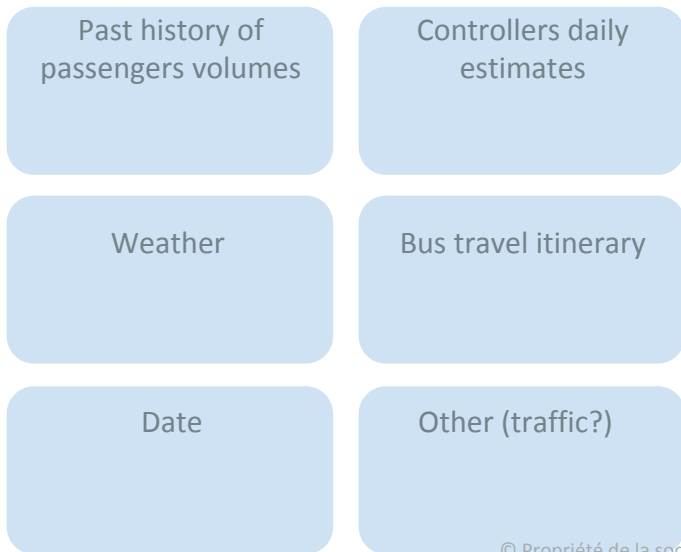
# HOW IT WORKS...

## HUPI ARTIFICIAL INTELLIGENCE ENGINE

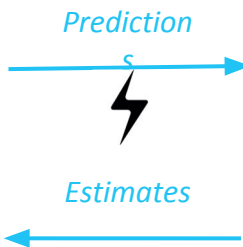
Machine Learning Algorithm



Data Sources

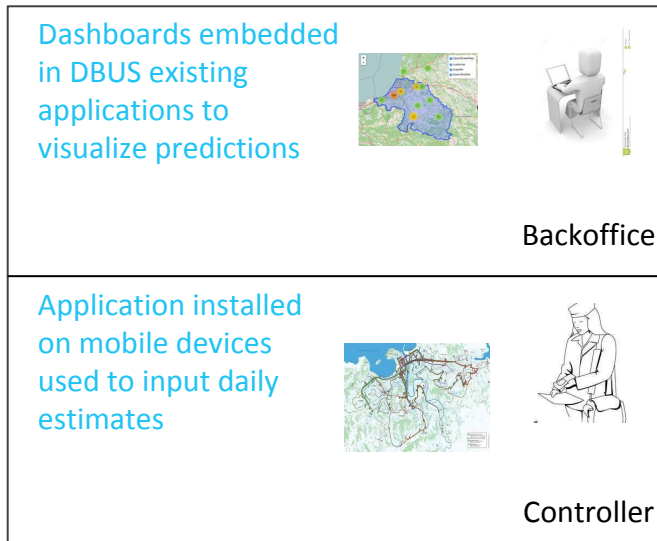


Internet

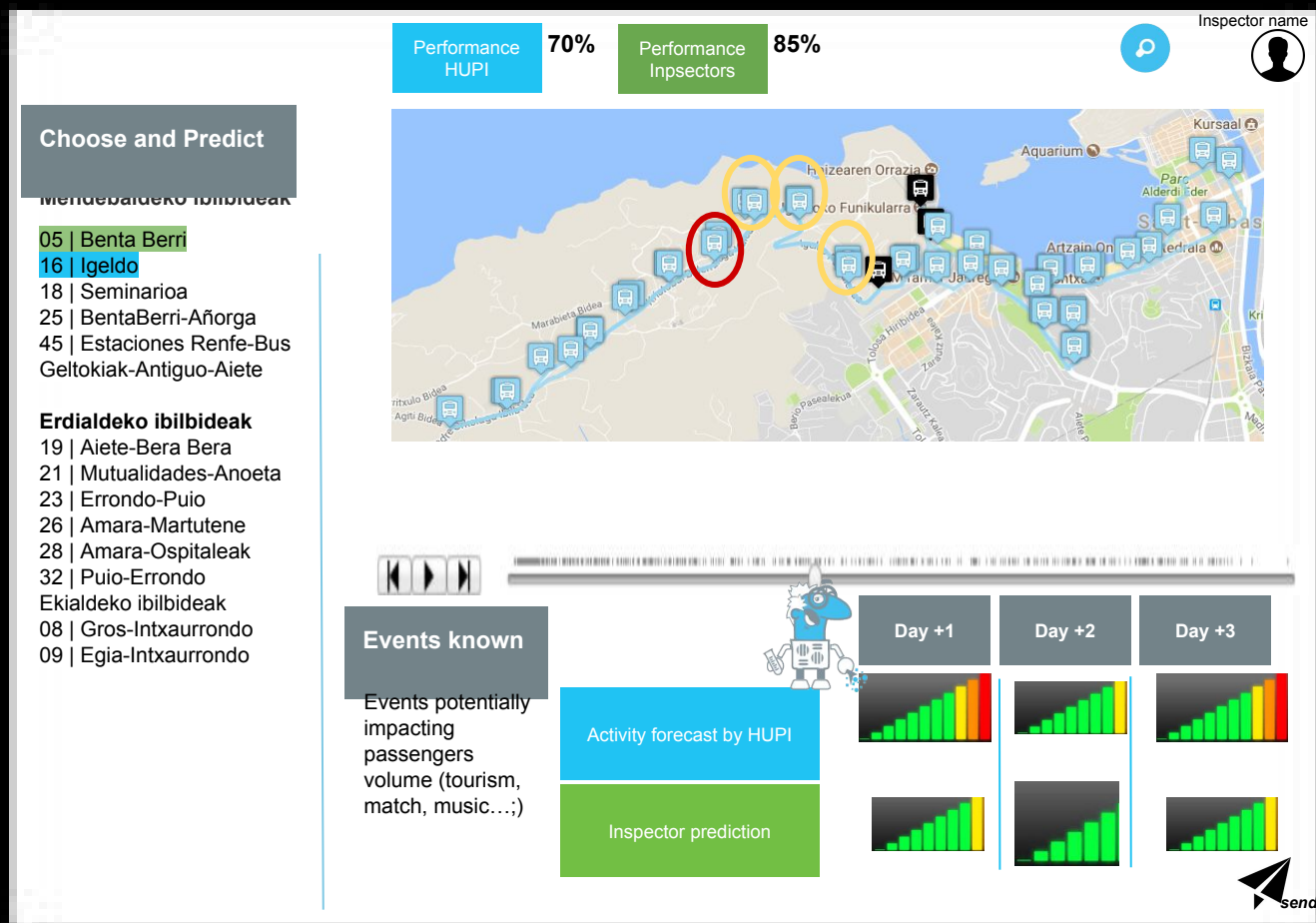


Real Time

## DBUS APPLICATION



# Example Mobile Applications



Inspectors will have a tool to provide their opinion and compare it to HUPI's prediction for the next 3 days  
*Important : the proposed design was made only to explain the process*



# FOCUS ON GEOSCIENCE

*Our Partners :*



## FOCUS ON GEOSCIENCE



### MÉTIER

Projet européen dans les Sciences Environnementales, initié et porté par le laboratoire CNRS GEODE « Université de Toulouse »,

en collaboration

- avec l'EISTI Pau,
- 2 universités européennes (Italie, Belgique),
- 2 universités Vietnamiennes
- 2 universités Thaïlandaises.

### BESOINS

Développer la recherche sur le cloud computing dans les sciences environnementales

Promouvoir son enseignement dans les pays partenaires d'Asie du Sud-Est:

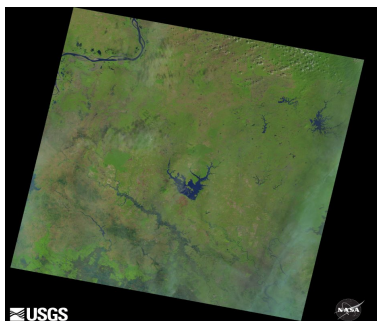
Sujets: Érosion, Pollutions atmosphériques urbaines en Asie du Sud-Est, Fonte du pergélisol, Systèmes d'alertes face aux risques environnementaux, Modélisation prospective des pratiques socio-spatiales et de l'occupation du sol

### SOLUTIONS

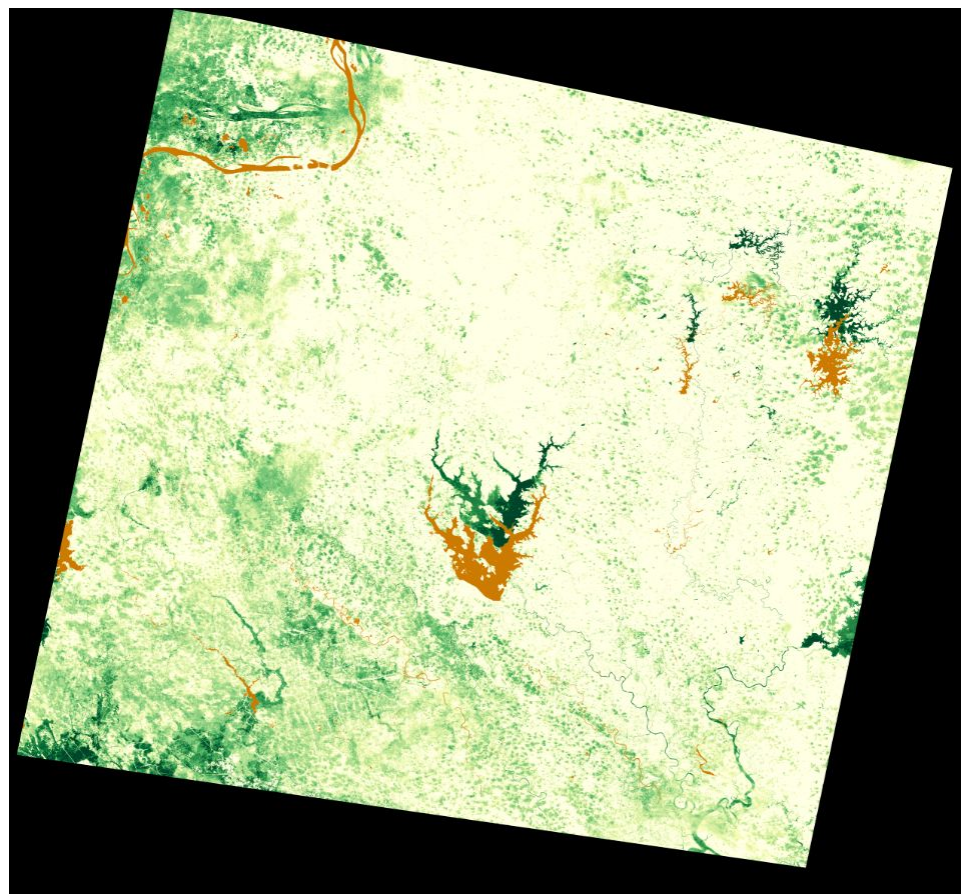
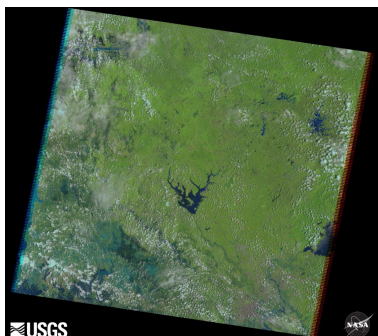
À l'oral

# COMPUTATION OF THE NDVI INDICATOR

Natural Color Image  
31-12-2017



Natural Color Image  
15-09-2007



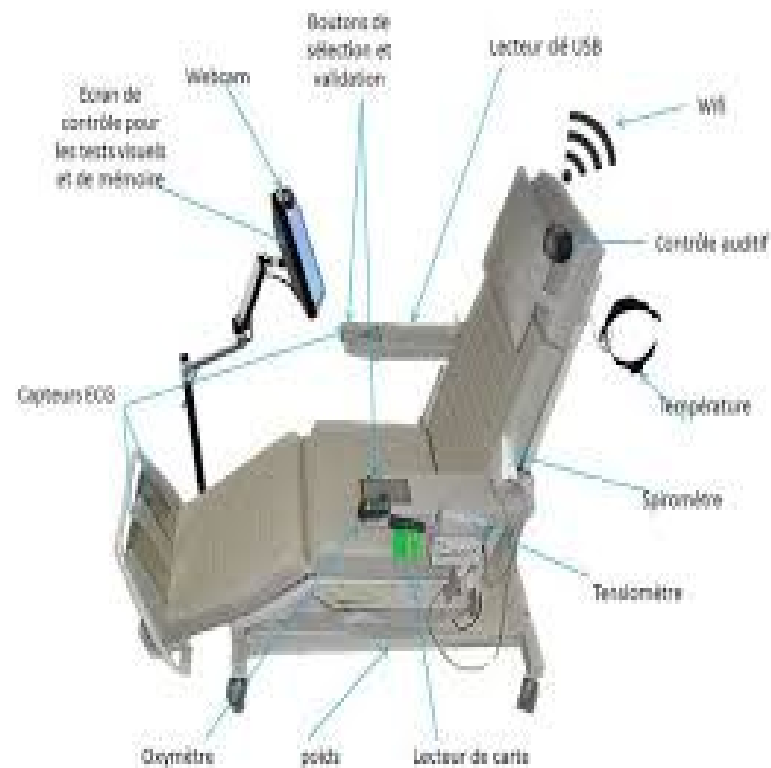
c. CASE STUDY



# CASE STUDY: SOKOA

Manufacturer based in Hendaye  
 French Leader of Office Chairs  
 (Turnover group > 100MM€+)

Plans to launch a “connected” chair



# OBJECTIVE

In groups of 2-4, present the services that SOKOA could develop, leveraging Artificial Intelligence and Big Data:

Preparation: 30mn

Presentation: 10mn

## SOKOA : IDEA #1. IMPROVE USER COMFORT

**Build a recommendation engine leveraging data usage to improve the comfort of each individual user**

### Usage Data :

- Settings
- Posture
- Usage
- Job function

### Recommendations :

- New settings
- Physical exercises
- Replace used components

# SOKOA : IDEA #2. MANAGE AND OPTIMIZE SPACE



QUESTIONS ?



# PROGRAM

## DAY 1 : Business Applications

- a. General Introduction : The New “Black Gold”
- b. Big Data for All !!!! - Examples of Real Applications
- c. Case Study : SOKOA

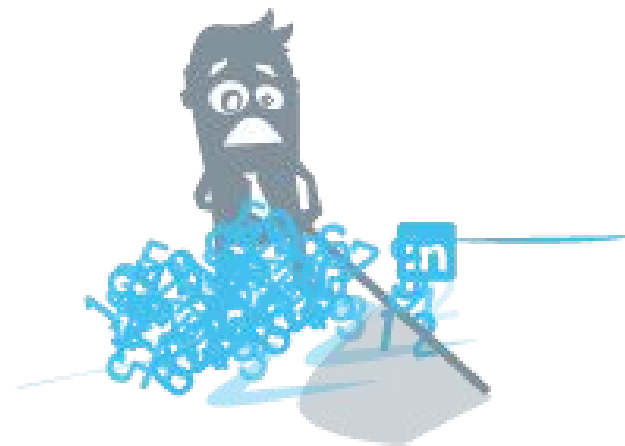
## DAY 2 : Artificial Intelligence and Big Data

- a. What is Changing?
- b. The Underlying Technologies
- c. Lab : See and use a Big Data Platform

## DAY 3 : How to Successfully Run a AI/Big Data Project

- a. The 3 “Pillars”
- b. Project : “Virtual Salesman” Marketing Strategy and Sales Pitch

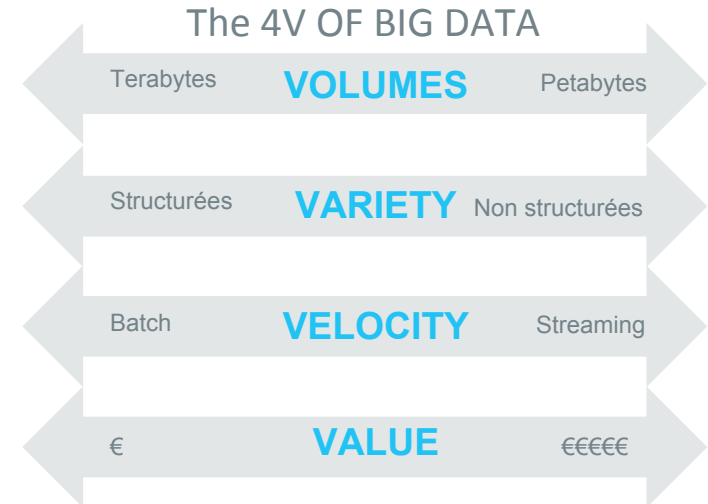
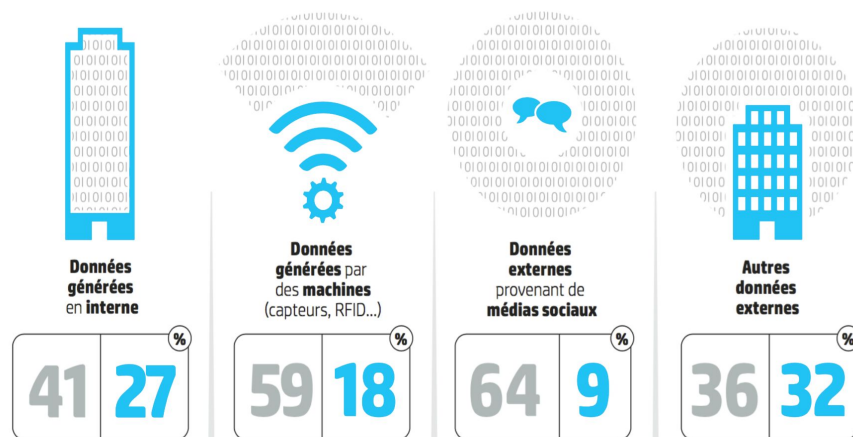
## a. WHAT IS CHANGING ?



# THE EXPLOSION OF DIGITAL DATA



## Importance de la mise en place d'outils Big Data autour des données



- Multiply sources and types of data
- Understand the environment more precisely
- Focus on data flows and not on stocks

## Une nouvelle conception du temps réel

**65%**



**L'analyse en temps réel est un enjeu majeur pour 65% des métiers**

**33%**

**Pour 1/3 des entreprises, aucun temps de latence acceptable pour l'analyse de leurs données**

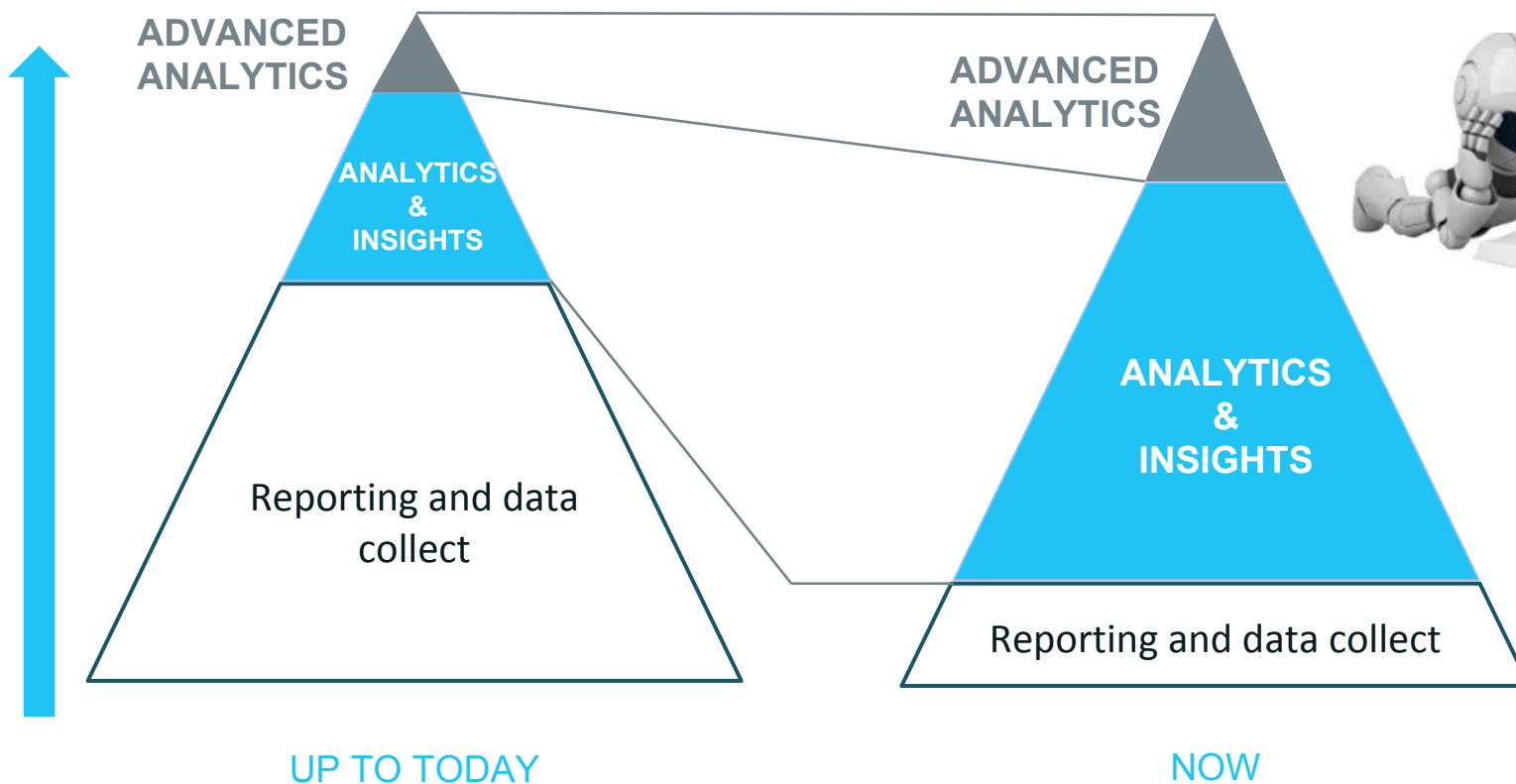
**40%**

**Pour 40% des entreprises, le temps de latence acceptable est inférieur à 10 minutes**

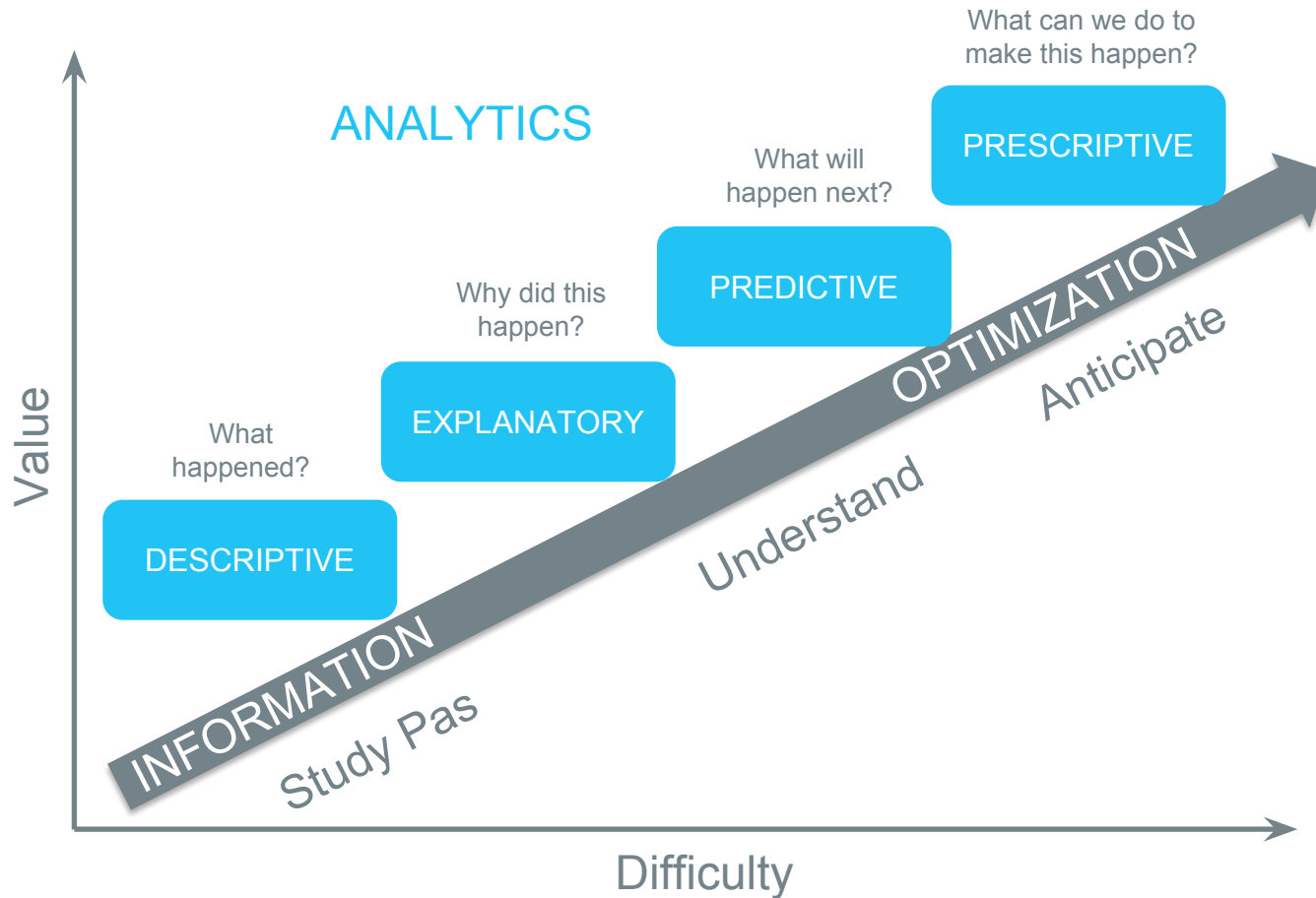
# ... TO ANALYZE BY DOING ADVANCED MATHEMATICS

VALUE  
CREATION

Machine Learning



# ... IN ORDER TO INFLUENCE THE FUTURE

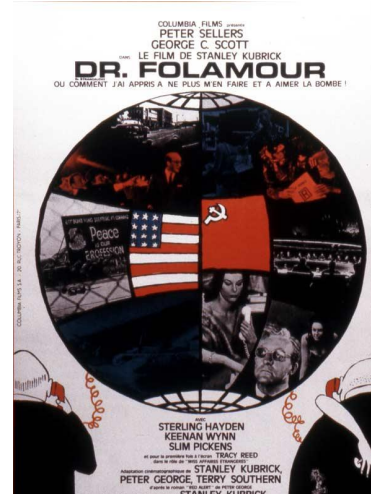


# | ... IN AUTOMATIC

## NUCLEAR DETERRENCE :

Extract of the movie Dr. Strangelove :  
1964, Stanley Kubrick :

- 4mn 37s



<https://www.youtube.com/watch?v=-mUCLHzWiJo>

## TRADING :

85% of the financial transactions in the US are today, entirely done by computers.

Thanks to complex algorithms, these computers are able to analyse in continuous market trades and react accordingly.

France Info October 2017



WITH GREAT POWERS...  
COME GREAT RESPONSIBILITIES...

UN GRAND POUVOIR  
IMPLIQUE DE GRANDES  
RESPONSABILITÉS



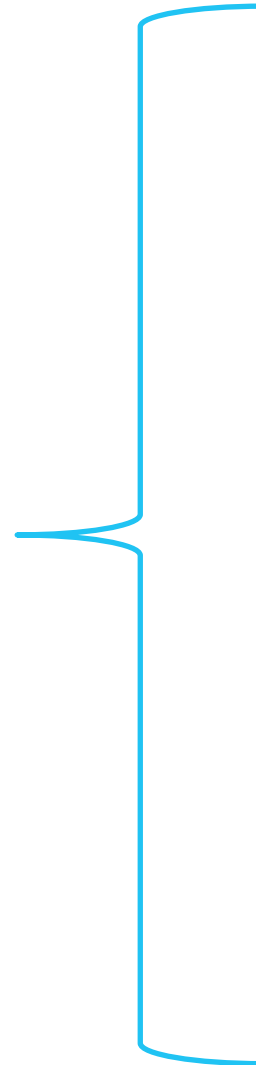
## ETHICAL CHALLENGES RELATED TO BIG DATA

A discussion by itself...

Juste a few words today...



# IN SUMMARY, BIG DATA = 4 SIMULTANEOUS ANSWERS



Data  
Everywhere

Real Time

Advanced  
Analytics

Automation



# SENTIMENT ANALYSIS



*CONNECT TO:*

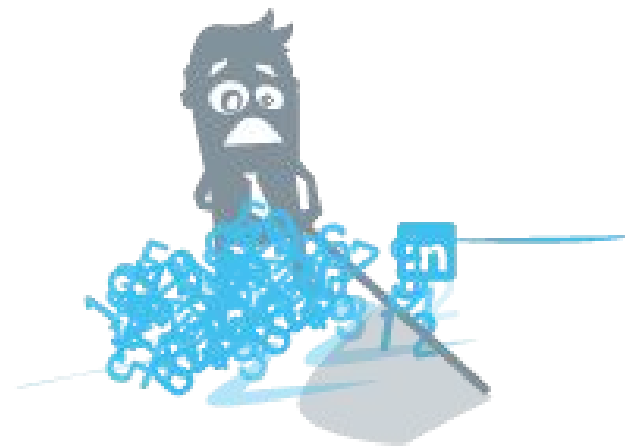
*<http://analysedesentiments.hupi.io>*

*AND TEACH THE MACHINE*

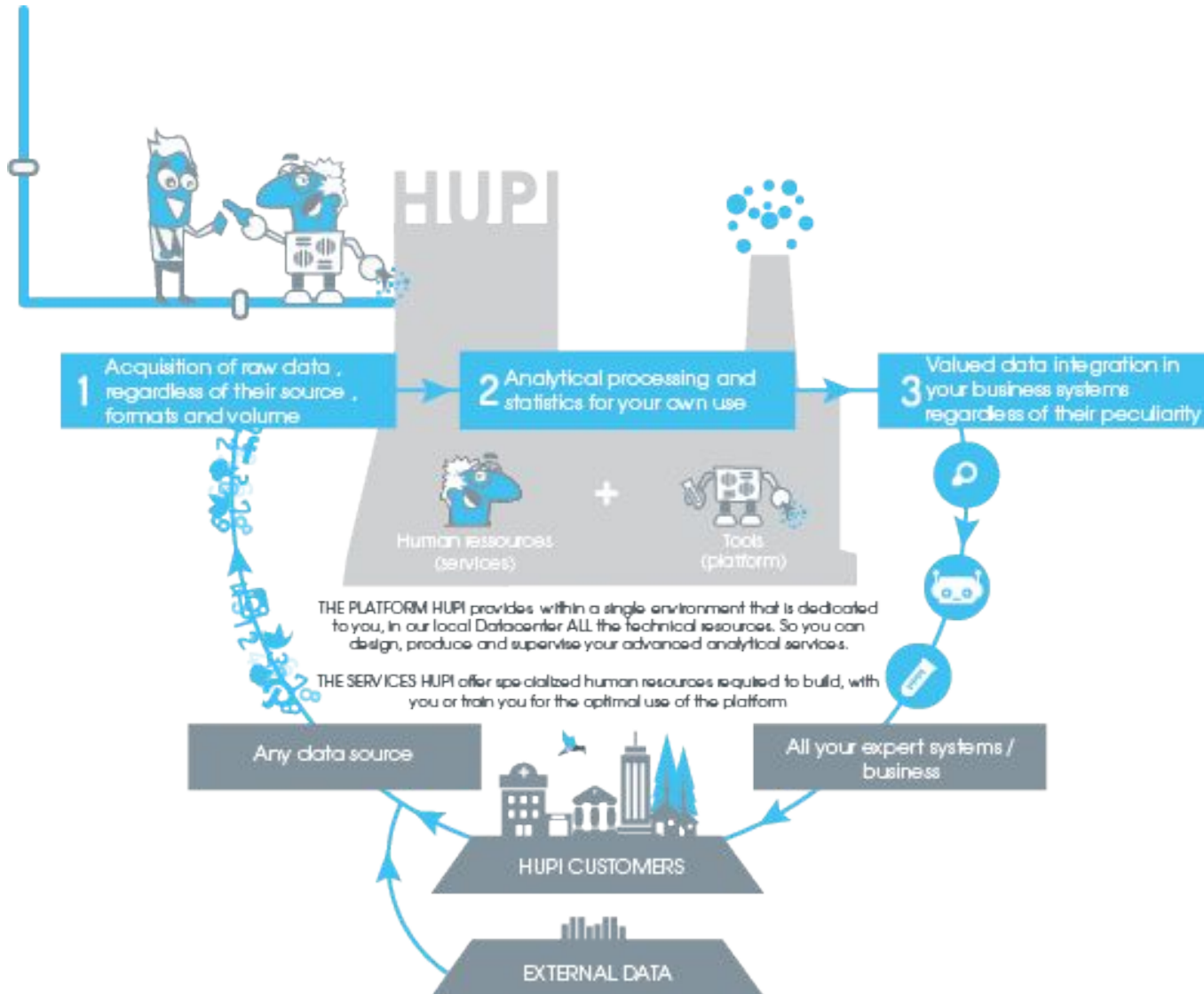
QUESTIONS ?



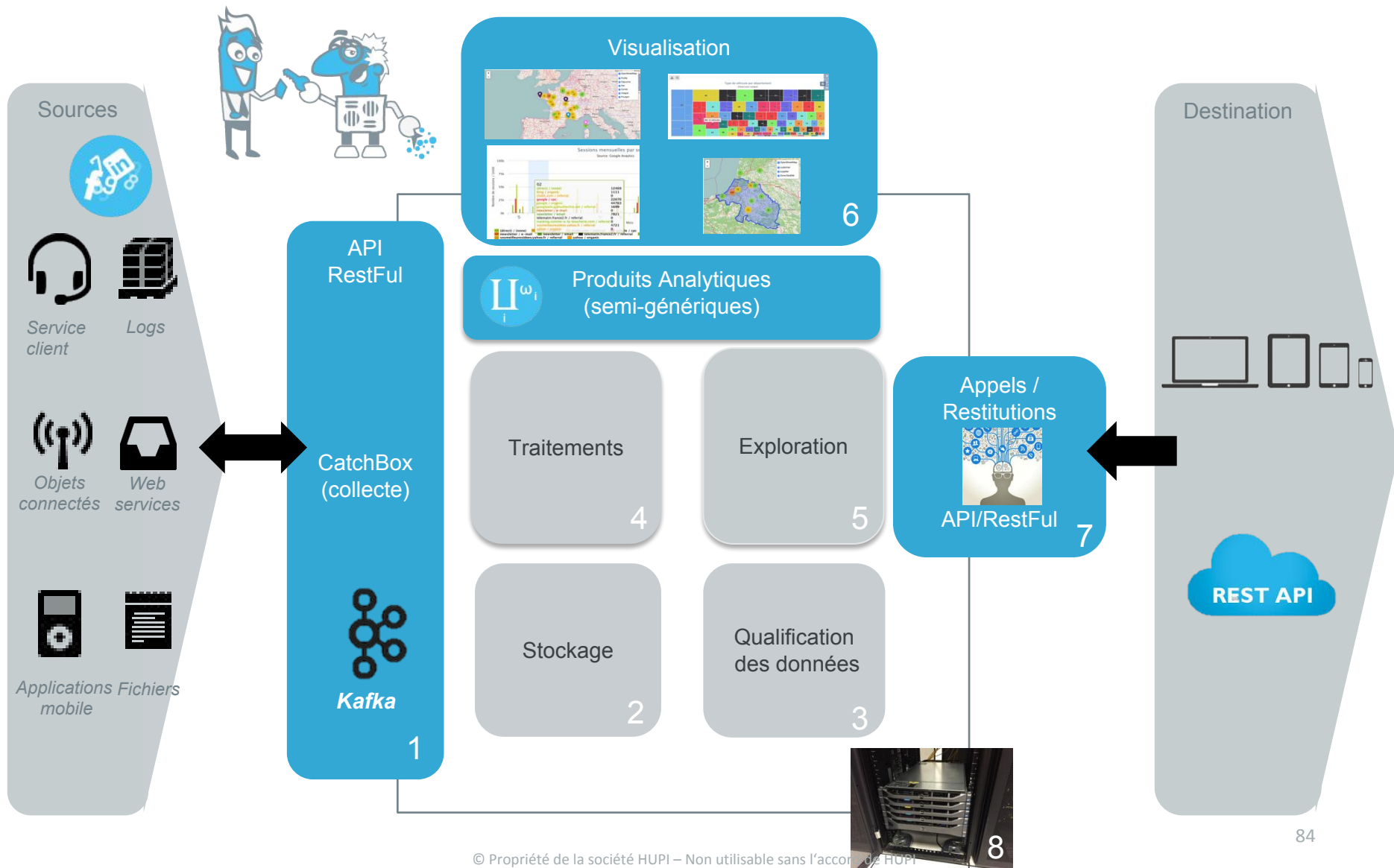
## b. TECHNICAL SOLUTIONS



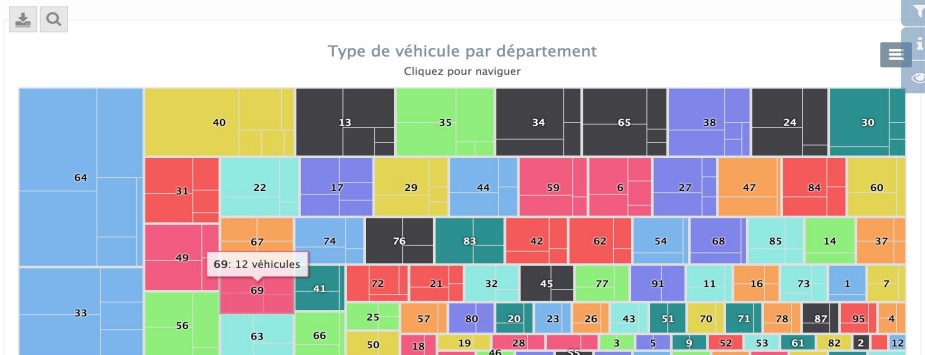
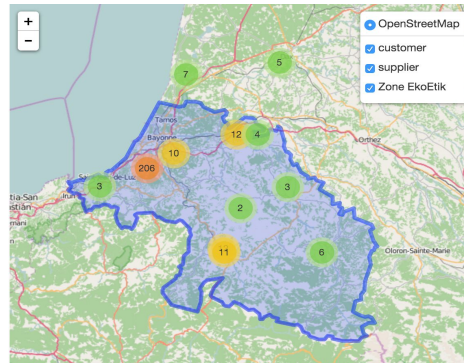
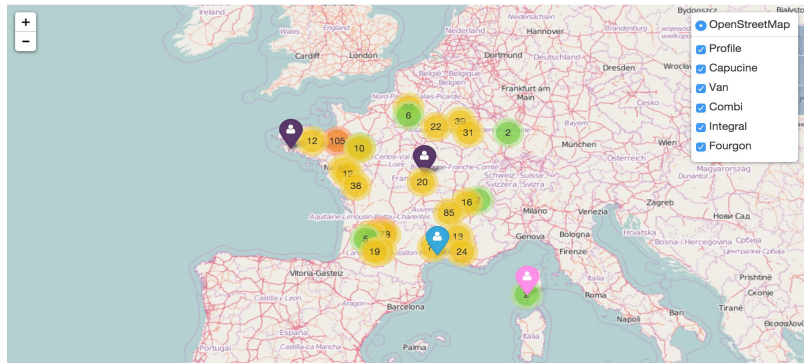
# HUPI SOFTWARE : A COMPLETE BIG DATA PLATFORM



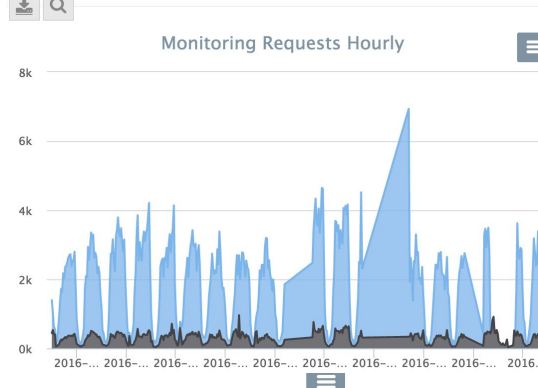
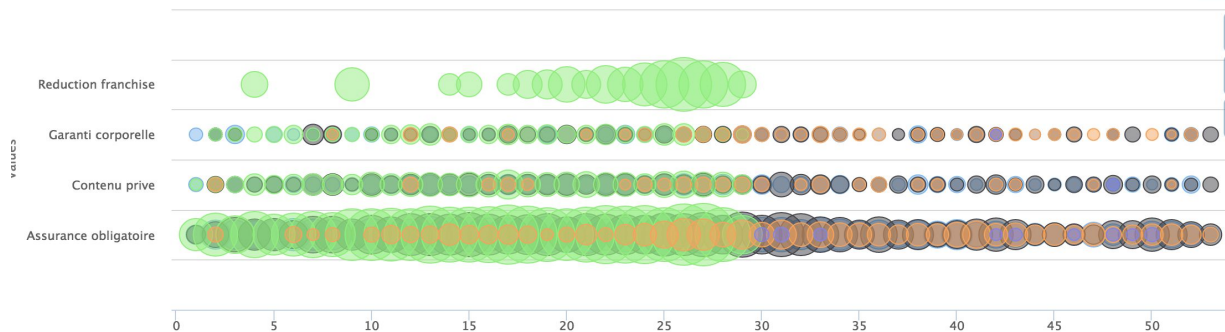
# ARCHITECTURE OF A BIG DATA PLATFORM



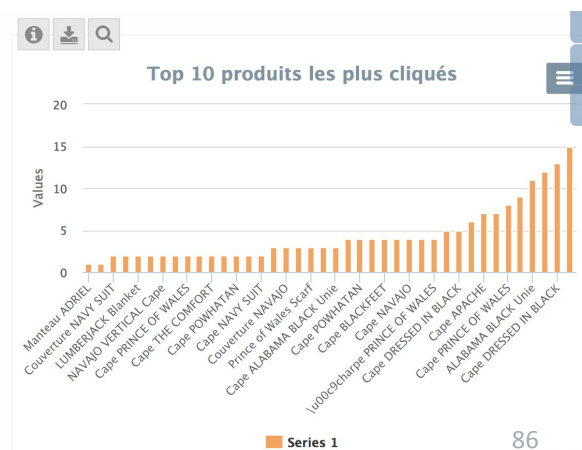
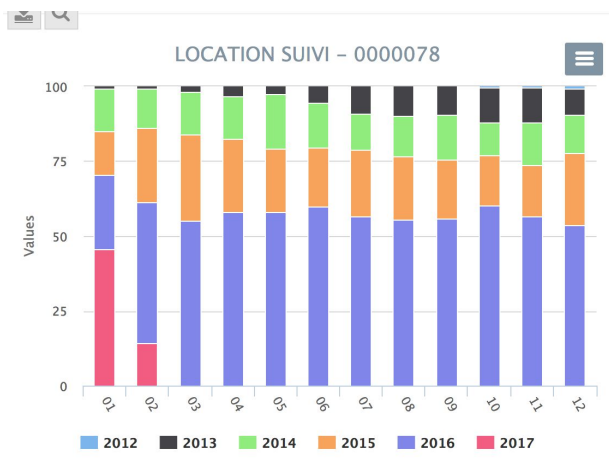
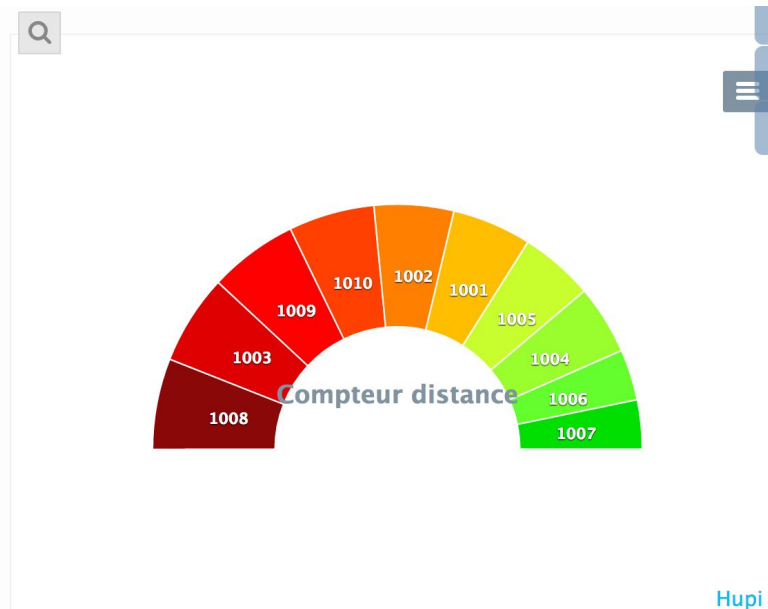
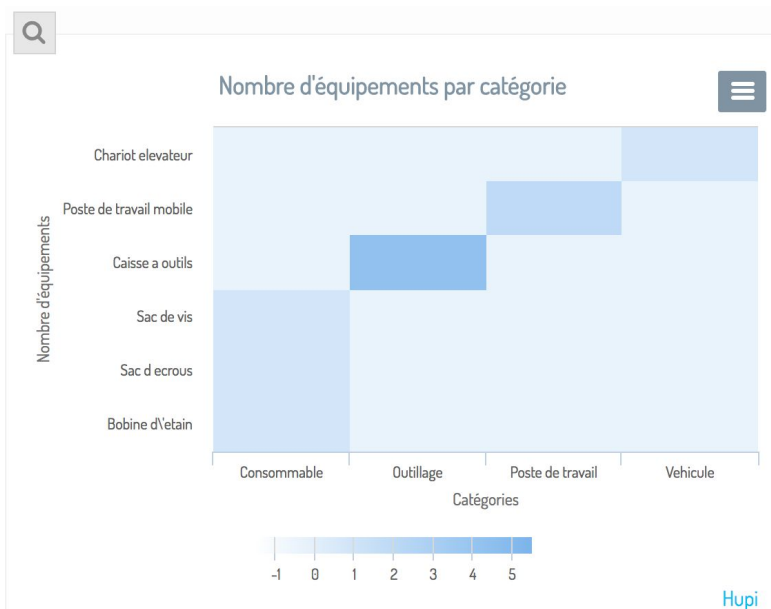
# VISUALIZATION SAMPLES



Chiffre d'affaires lié à l'assurance



# VISUALIZATION SAMPLES



# TO TRACK YOUR EQUIPMENT

- ✕ Stratégie ▾
- 📄 Locataires ▾
- 🏠 Propriétaires ▾
- 🚗 Vehicules ▾
- 📅 Locations ▾
- € Paiements ▾
- € Commercial ▾
- 🔄 Service client ▾
- 📢 Marketing ▾
- 📄 Export Données ▾
- 📄 Recommandations ▾
- ☰ Suivi des erreurs ▾
- ⚡ Performance ▾
- G+ Google Analytics - Blog ▾
- G Google Analytics - ecommerce ▾

## Stocks → Etat des consommables

+ Afficher les critères

↓ **etain**  
Niveau correct

↑ **vis**  
Niveau très faible

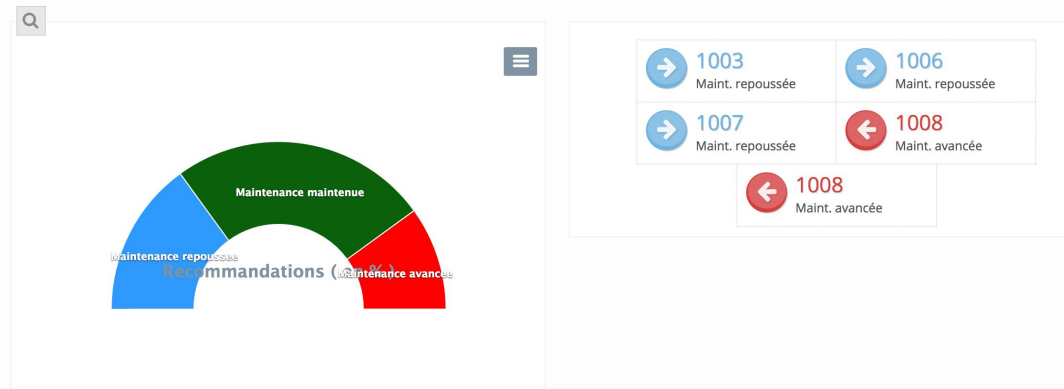
↑ **ecrous**  
Niveau faible



# TO ANTICIPATE CHANGES

## Maintenance équipements → Maintenance et recommandations

+ Afficher les critères



### Maintenances prévues et recommandées

Tag_id	Equipement	Categorie	Type	Date_prevu	Date_predite
1004	Caisse a outils	Outillage	Verification exhaustive contenu	30-09-2016	30-09-2016
1006	Caisse a outils	Outillage	Verification exhaustive contenu	30-09-2016	03-10-2016
1008	Bobine d'etain	Consommable	Reapprovisionnement	30-09-2016	29-09-2016
1009	Poste de travail mobile	Poste de travail	Verification etat	03-10-2016	04-10-2016
1007	Chariot elevateur	Vehicule	Controle technique vehicule	07-10-2016	05-10-2016
1003	Caisse a outils	Outillage	Verification exhaustive contenu	30-09-2016	03-10-2016
1002	Sac de vis	Consommable	Reapprovisionnement	30-09-2016	30-09-2016
1010	Poste de travail mobile	Poste de travail	Verification etat	03-10-2016	03-10-2016
1001	Sac d ecrous	Consommable	Reapprovisionnement	30-09-2016	30-09-2016
1005	Caisse a outils	Outillage	Verification exhaustive contenu	30-09-2016	30-09-2016

# PROJECT : VIRTUAL SALESMAN

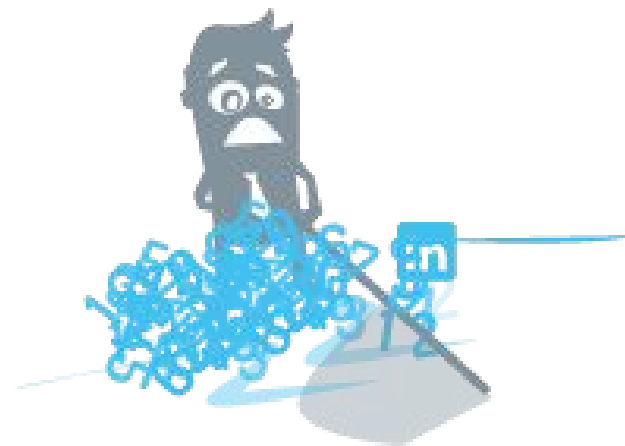
Do you have any question?



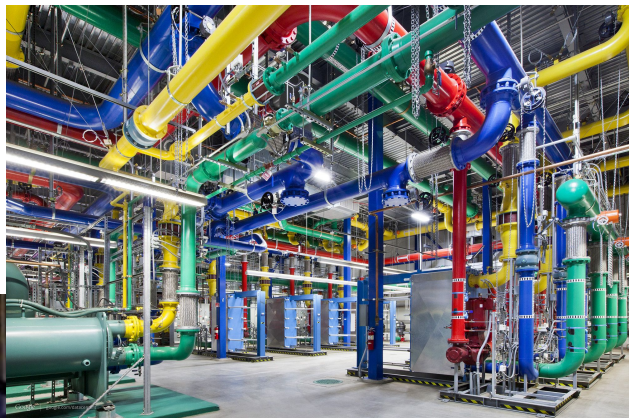
QUESTIONS ?



## C. LAB



# DATA CENTER



*facebook 2nd datacenter*



*Google DataCenter*

# HADOOP : Big Data Technology



## What is Big Data - Intelligence Artificial

Big data is **massive and messy**, and it's coming at you **uncontrolled**. Data are gathered to be analyzed to **discover patterns and correlations** that could not be initially apparent, but might be useful in **making business decisions** in an organization. These data are often personal data, which are useful from a marketing viewpoint to **understand the desires and demands of potential customers** and in **analyzing and predicting their buying tendencies**.

# HADOOP : Big Data Technology



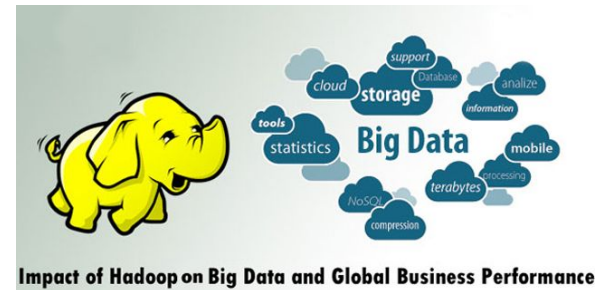
## Breaking through Silos

Silos are a result of **hierarchies of the organization**, which require organizing people into economically effective groups.

**Data silos become a barrier** that impedes decision-making and organizational performance. Enterprises are facing many challenges to glean insight with Big Data Analytics that trapped in the data silos exist across business operations. Through the effective handling of big data can stymie data silos and the **enterprise can leverage available data into emerging customer trends or market shifts for insights and productivity.**

# Excercice

## How you will do it ?



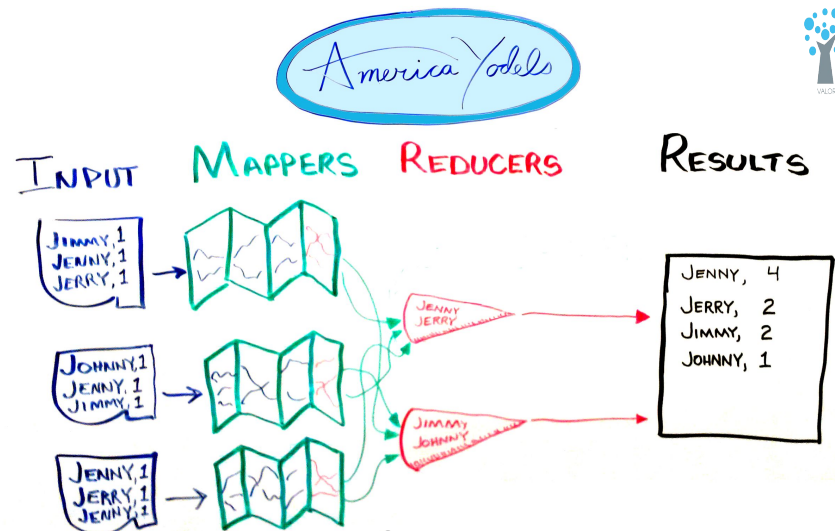
Give me different solutions to count the occurrence of each words that appears in this document:

To be, or not to be, that is the question:  
Whether 'tis nobler in the mind to suffer  
The slings **and** arrows of outrageous fortune,  
Or to take Arms against a Sea of troubles,  
**And** by opposing end them: to die, to sleep  
No more; **and** by a sleep, to say we end  
the heart-ache, **and** the thousand natural shocks  
that Flesh is heir to? 'Tis a consummation  
devoutly to be wished. To die, to sleep,  
To sleep, perchance to Dream; aye, there's the rub,  
for in that sleep of death, what dreams may come,  
when we have shuffled off this mortal coil,  
must give us pause. There's the respect  
that makes Calamity of so long life:  
For who would bear the Whips **and** Scorns of time,  
the Oppressor's wrong, the *proud* man's Contumely, [F: *poor*]  
the pangs of *despised* Love, the Law's delay, [F: *disprized*]  
the insolence of Office, **and** the spurns  
that patient merit of the unworthy takes,  
when he himself might his Quietus make  
with a bare Bodkin? Who would Fardels bear, [F: *these*  
*Fardels*]

to grunt **and** sweat under a weary life,  
but that the dread of something after death,  
the undiscovered country, from whose bourn  
no traveller returns, puzzles the will,  
**and** makes us rather bear those ills we have,  
than fly to others that we know not of.  
Thus conscience does make cowards of us all,  
**and** thus the native hue of Resolution  
Is sicklied o'er, with the pale cast of Thought,  
**And** enterprises of great *pitch* **and** moment, [F:  
*pith*]  
with this regard their Currents turn *awry*, [F:  
*away*]  
**And** lose the name of Action. Soft you now,  
The fair Ophelia? Nymph, in thy Orisons  
Be all my sins remember'd

AND: 12  
.....

# HADOOP : Big Data Technology



In summary: Hadoop = Map Reduce + HDFS + ecosystem tools.

**Map Reduce** aims to execute batch processing on a distributed system.

The idea is that we manage data that is a priori unstructured, so we have a step of extracting and processing the relevant data.

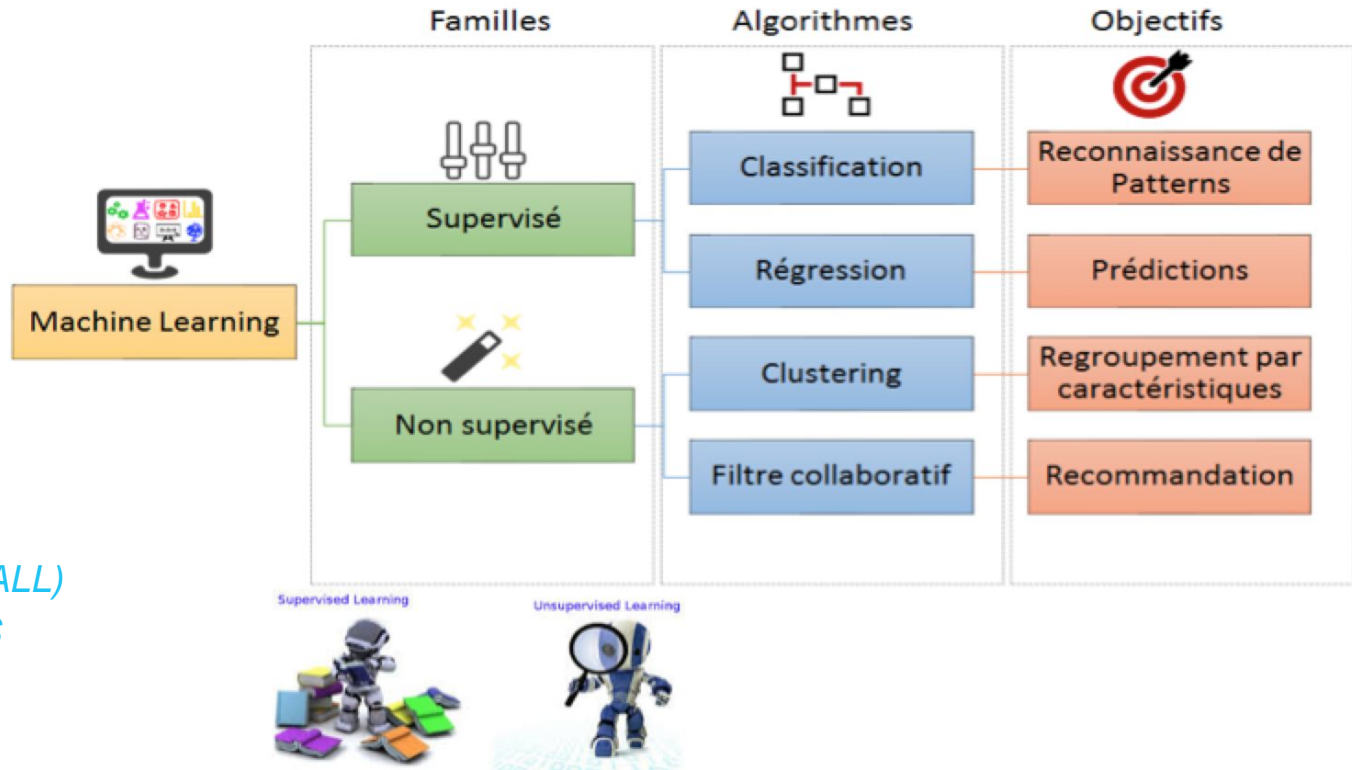
What is called **Map**.

It is to read information in a file, to extract the part which concerns us, and to present it in the form of key value. Finally, we obtain a loooooongue list, distributed, of key - value pairs.

A step of processing the relevant data, which bears the sweet name of Reduce. As much as the Map brings us back lists of key - value pairs, as much the reduce will treat, and aggregate, these values.



# FOCUS: MACHINE LEARNING : DETERMINISM vs. MODELING



*DECIDE ALSO (ABOVE ALL)  
BASED ON STATISTICS*

AND

*NO MORE (NECESSARILY)  
BASED ON RULES*



## Les Technologies :

# Apache Spark



Scaling > 10 000 computers

- Stockage et traitement de **exabytes** de données

Vers l'infiniment grand			
Préfixe			Peu utilisée
<u>Yotta</u>	X 1 000 000 000 000 000 000 000 000 000 000 000	<b>10<sup>24</sup></b>	Quadrillion
<u>Zetta</u>	X 1 000 000 000 000 000 000 000 000 000	<b>10<sup>21</sup></b>	Trilliard
<u>Exa</u>	X 1 000 000 000 000 000 000 000	<b>10<sup>18</sup></b>	Trillion
<u>Péta</u>	X 1 000 000 000 000 000 000	<b>10<sup>15</sup></b>	Billiard
<u>Téra</u>	X 1 000 000 000 000 000	<b>10<sup>12</sup></b>	Billion
<u>Giga</u>	X 1 000 000 000	<b>10<sup>9</sup></b>	Milliard
<u>Méga</u>	X 1 000 000	<b>10<sup>6</sup></b>	Million
<u>Kilo</u>	X 1 000	<b>10<sup>3</sup></b>	Mille
<u>Hecto</u>	X 100	<b>10<sup>2</sup></b>	Cent
<u>Déca</u>	X 10	<b>10<sup>1</sup></b>	Dix

- Tolérance aux pannes lors des échecs (relance automatique)
- 2 tiers des serveurs peuvent “crasher” et pourtant la requête sera effectuée
- Indépendance du matériel et des logiciels

# Notebooks

- Connectez-vous à <https://ecoles.hupi.io>
- créer votre compte
- je vous autorise
- créé votre compte exper
- Ouvrez un notebook (de préférence avec Chrome ou FireFox)

Ouvrez le dossier UPPA - Formation 2017

# présentation des outils notebooks

# démo d'un calcul d'indicateurs

# démo d'un calcul de similarité

# PROGRAM

## DAY 1 : Business Applications

- a. General Introduction : The New “Black Gold”
- b. Big Data for All !!!! - Examples of Real Applications
- c. Case Study : SOKOA

## DAY 2 : Artificial Intelligence and Big Data

- a. What is Changing?
- b. The Underlying Technologies
- c. Lab : See and use a Big Data Platform

## DAY 3 : How to Successfully Run a AI/Big Data Project

- a. The 3 “Pillars”
- b. Project : “Virtual Salesman” Marketing Strategy and Sales Pitch

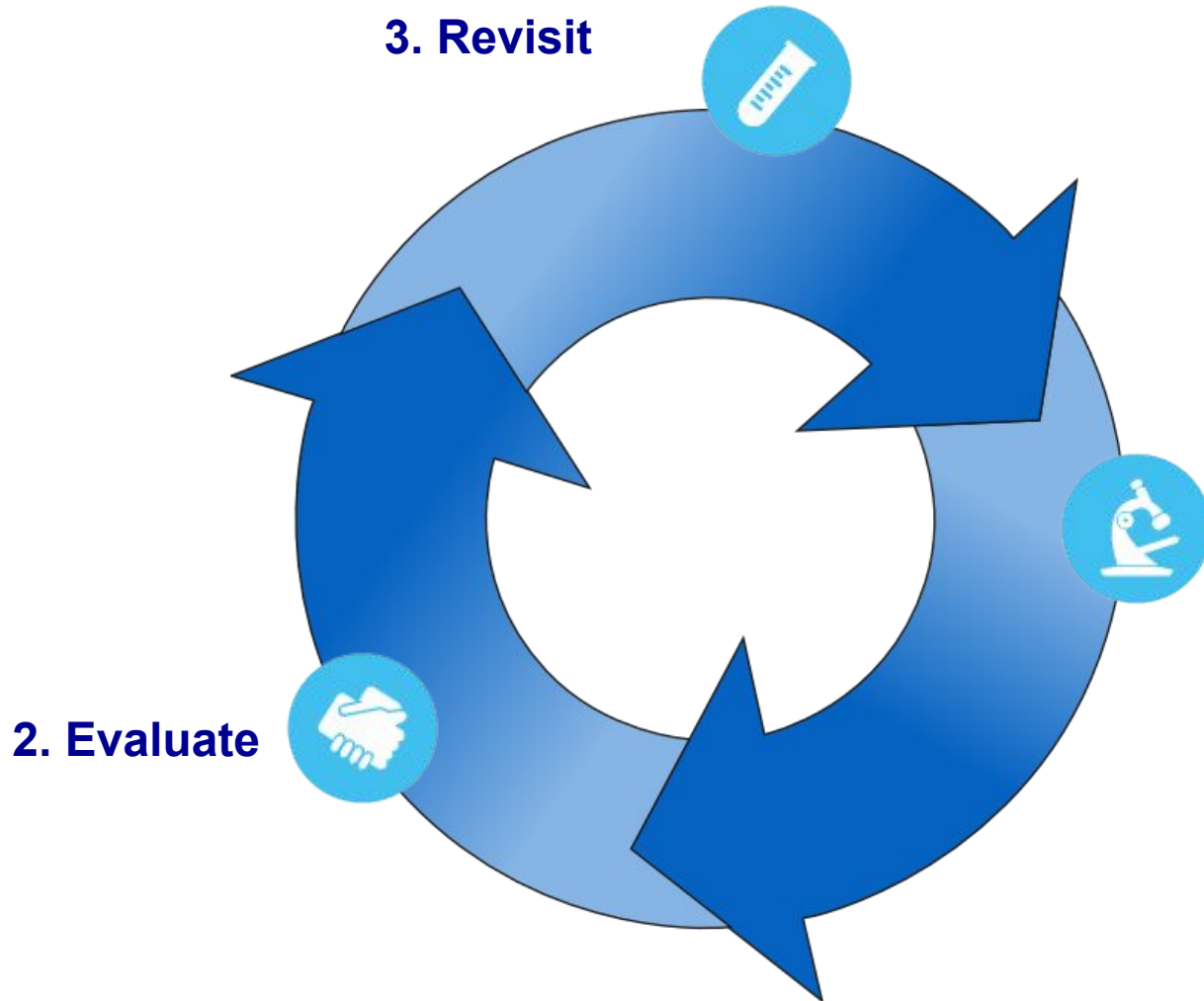
# TO REFRESH OUR MEMORY...

TO START A BIG DATA PROJECT, IT IS NECESSARY TO	TRUE
... have a lot of data ...	
... have an IT team in my company...	
... love maths...	
... be very rich...	
... be very patient ...	
... know exactly what clients want	
... know perfectly the state of its own data...	
... be in specific businesses : bank, insurance...	
... be a very large company ...	

## a. THE 3 PILLARS

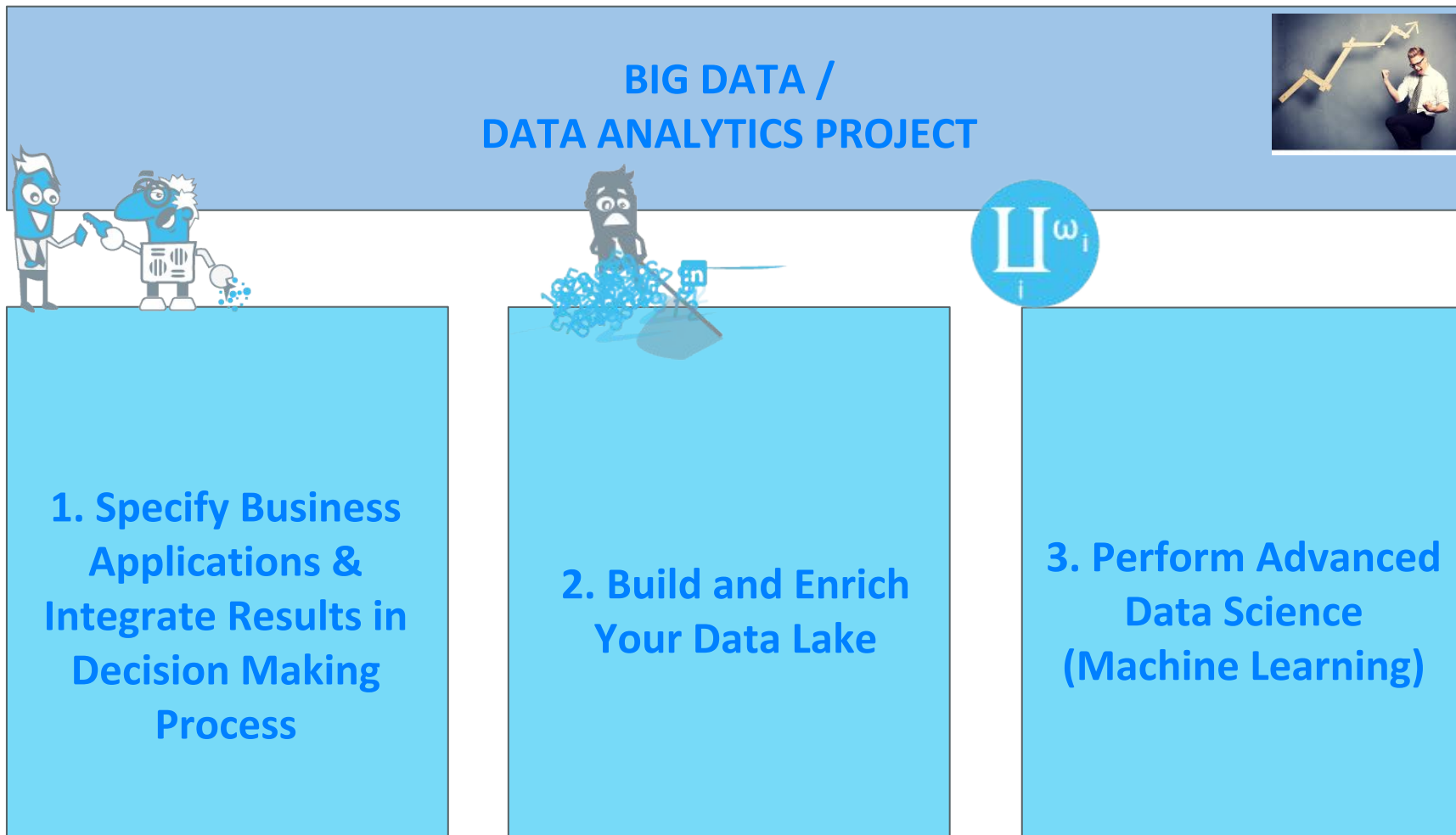


# TRY TO LEARN, THEN DECIDE...

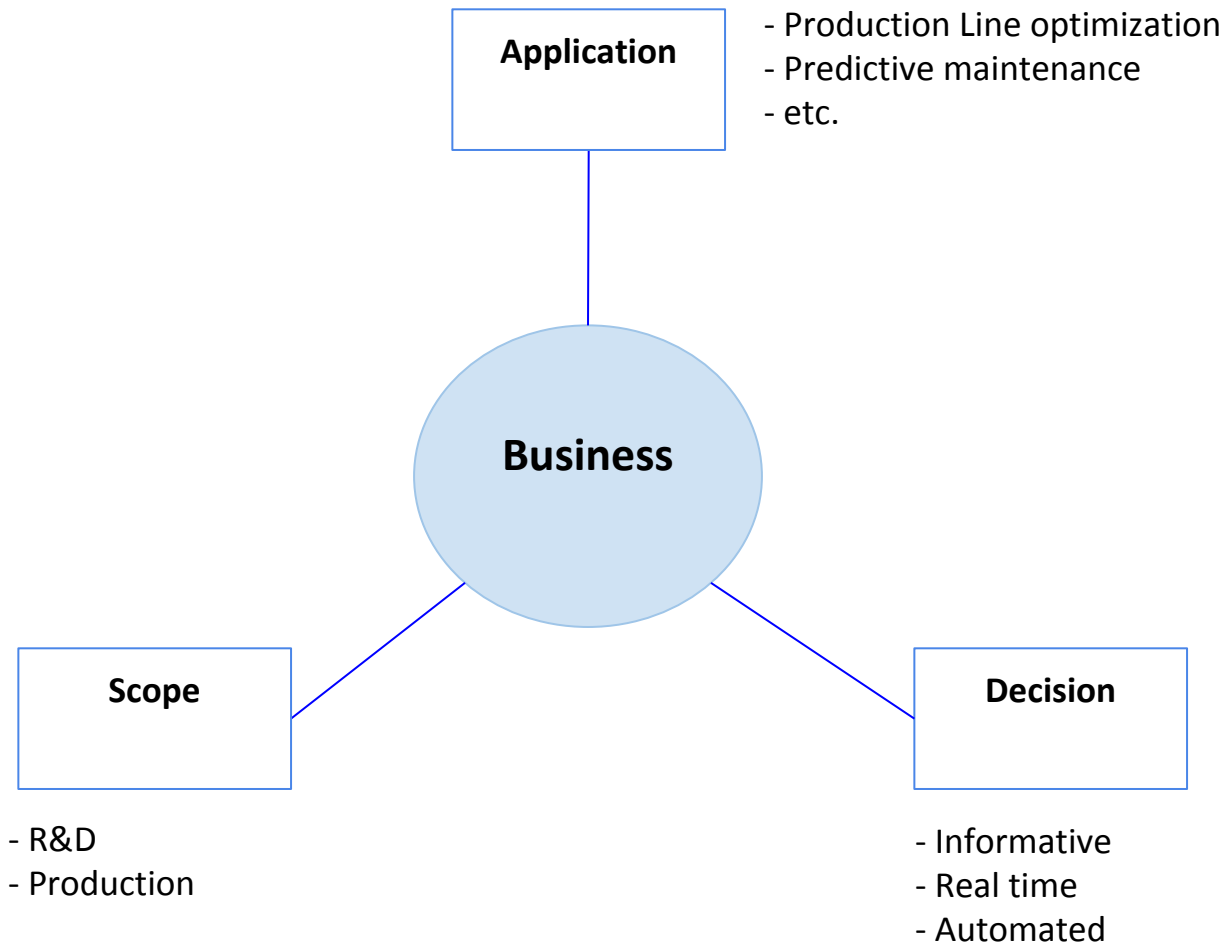


## 1. Experiment

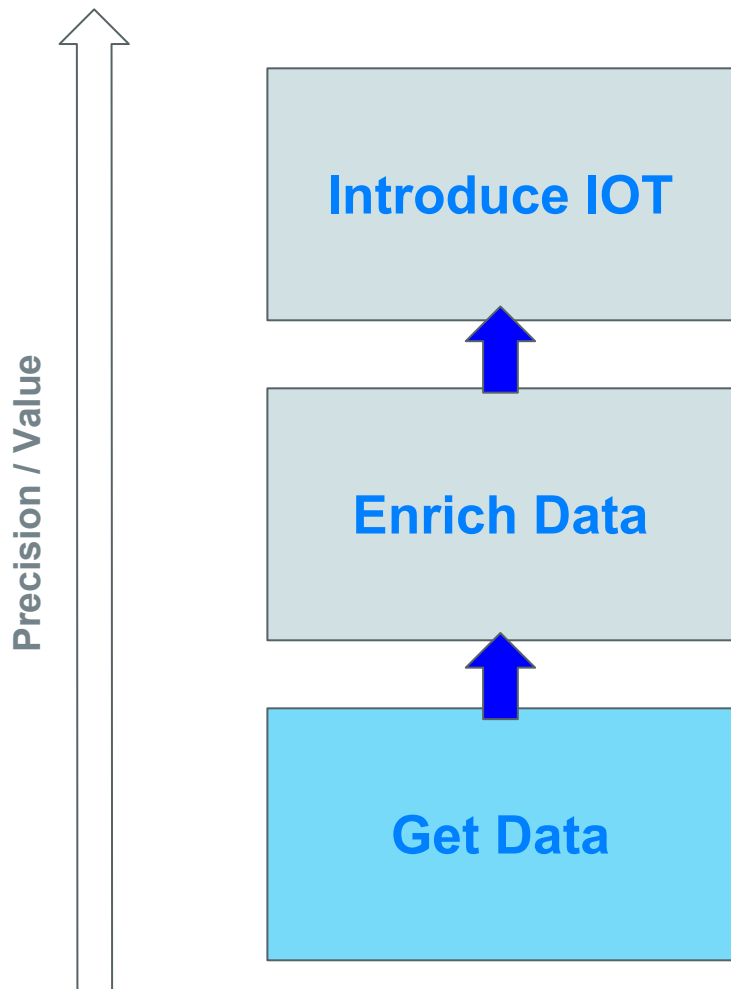
# THE 3 PILLARS OF A SUCCESSFUL “BIG DATA”/DATA ANALYTICS PROJECT



# 1. SPECIFY BUSINESS APPLICATIONS & INTEGRATE RESULTS IN DECISION MAKING PROCESS



## 2. BUILD AND ENRICH YOUR DATA LAKE

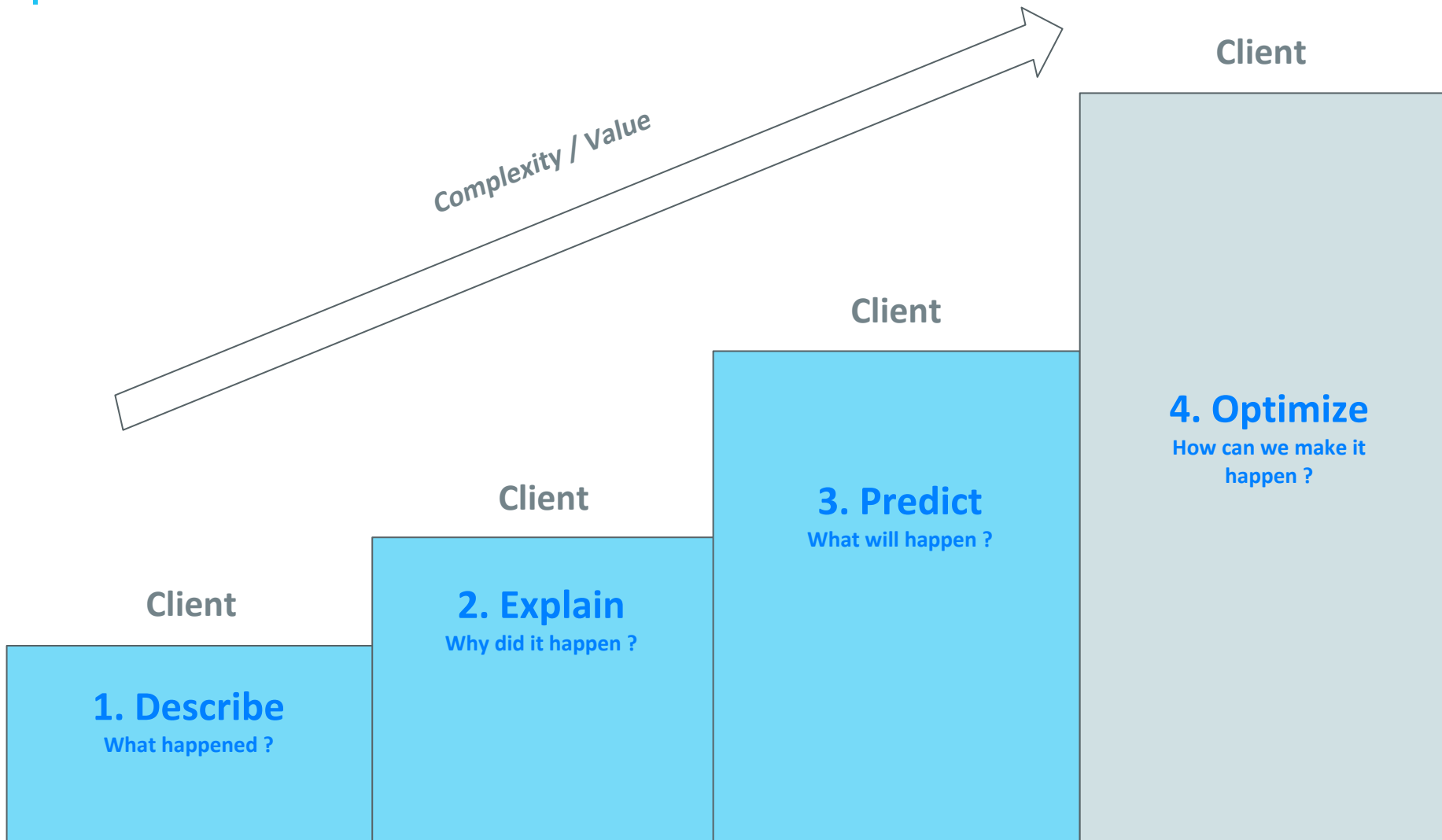


Introduce connected objects to measure and communicate : machines, control tools, etc.

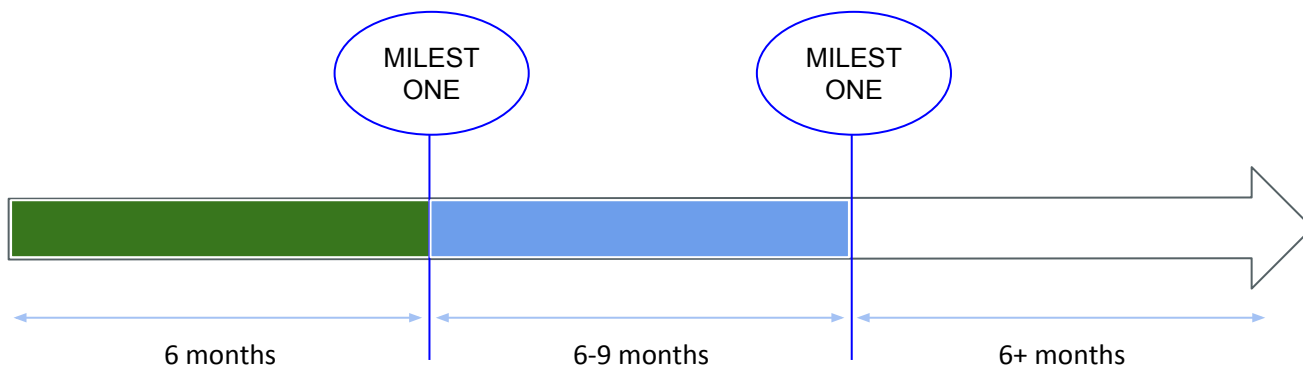
Add new data sources and unstructured data to your data lake : financial, HR, etc.

Digitalize your events, information and knowledge : production, quality, etc.

## 3. PERFORM ADVANCED SCIENCE

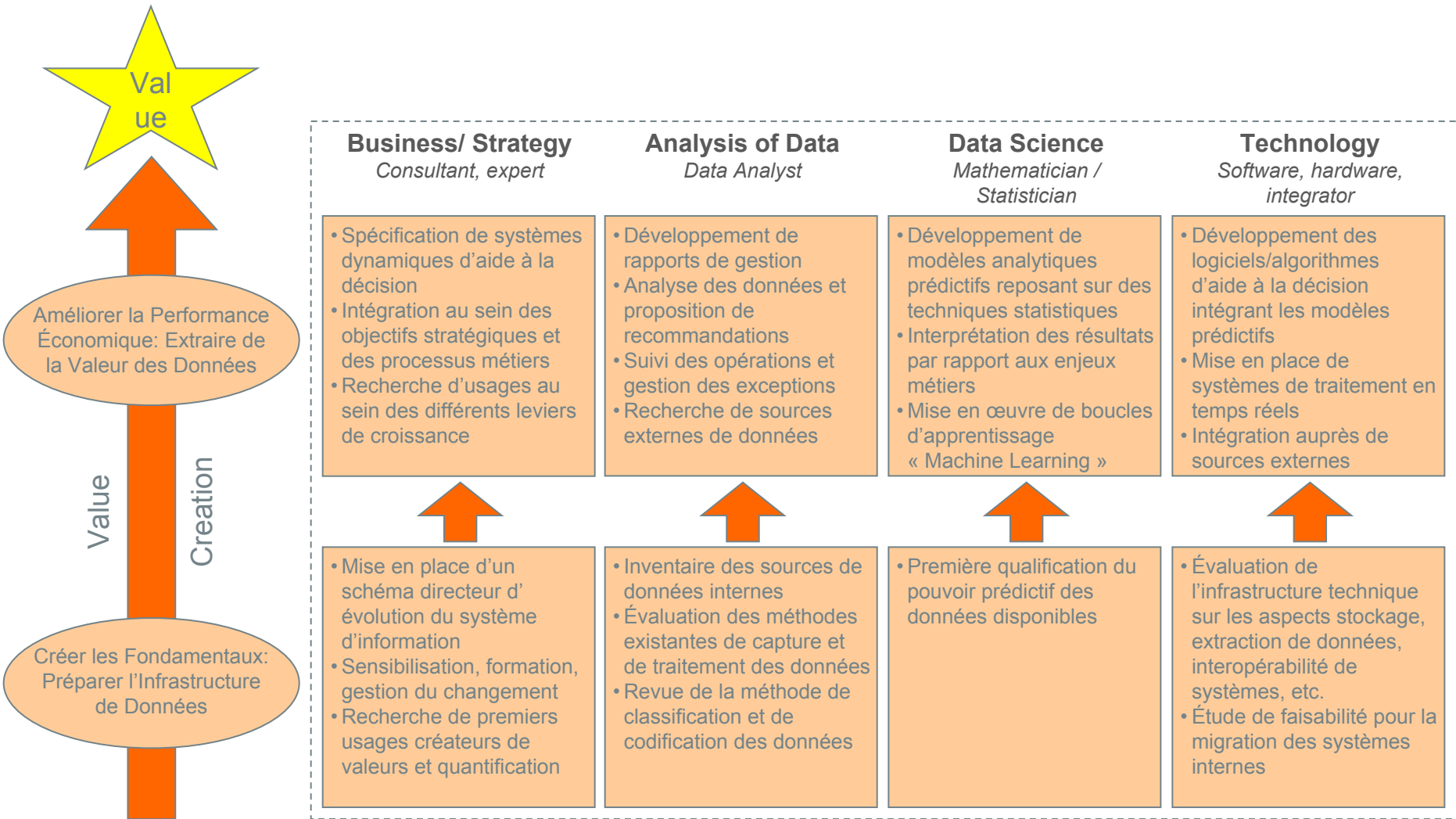


# POSSIBLE ROADMAP

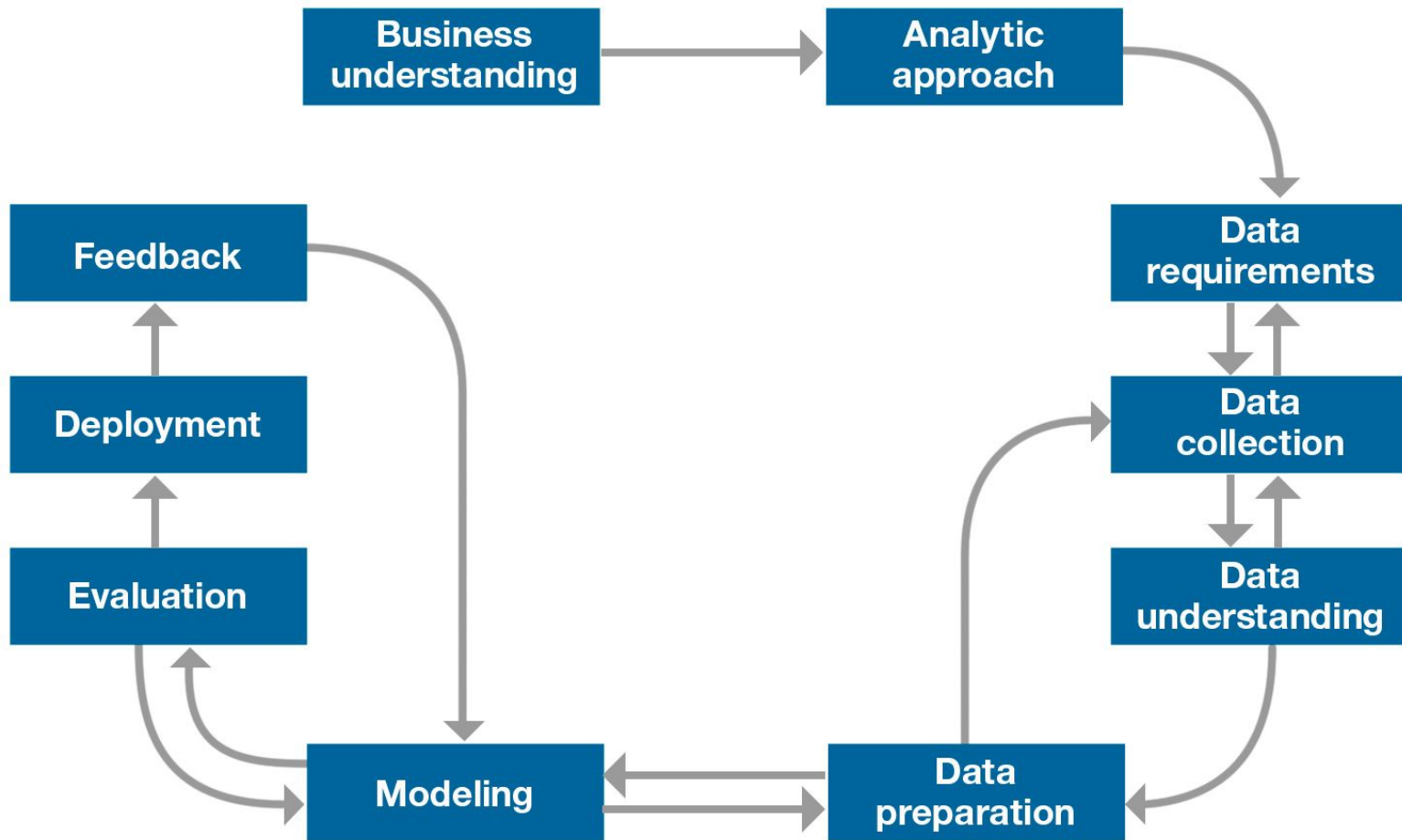


<b>1. BUSINESS APPLICATION</b>	Line Optimization R&D Informative Informative: results available on HUPI platform IHM	Line Optimization R&D Informative: results available on client dedicated IHM	Line Optimization + other R&D + production Informative + real time alerts
<b>2. DATA LAKE</b>	Get Data (production/quality) on a simple data acquisition interface	Enrich Data Lake with other data sources (TBD) and through an advanced IHM	Introduce IOTs
<b>3. DATA SCIENCE</b>	Describe Explain Predict on quality outputs	Describe Explain Predict on more outputs	Optimize for multiple outputs

# PROFILES REQUIRED TO PURSUE A BIG DATA PROJECT



# POSSIBLE ROADMAP



## b. THE “VIRTUAL SALESMAN”



## PROJECT : VIRTUAL SALESMAN

### Objective:

- develop your recommendations to design an efficient Marketing Strategy to scale-up the sales of “HUPI Virtual Salesman” product
- prepare a 5mn pitch, with responses to objections, to sell the Virtual Salesman to the e-commerce firm Scandivie.



Presentations will take place in groups of 2-4,  
on Friday October 27th afternoon

- Marketing Strategy: 30mn
- Sales Pitch: 5mn

# PROJECT : VIRTUAL SALESMAN

## HUPI's Approach and Strategy



QUESTIONS ?

