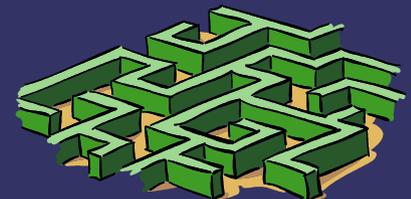


La production, le stockage et l'exploitation des données. Quels enjeux ?



ICAM

Union Régionale des Ingénieurs et
Scientifiques de Midi-Pyrénées



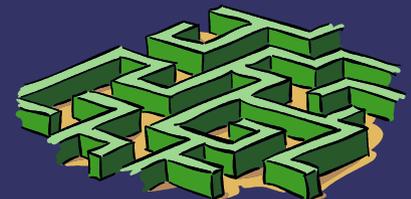
Sommaire

- 1 – La production de données, origine et impact du Big Data
- 2 – Le traitement et l'exploitation des données
- 3 – Les opportunités, les risques d'un monde ouvert
- 4 – Conclusion



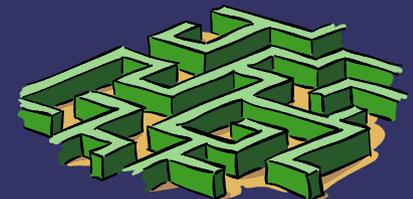
1

La production de données, origine et impact du Big Data



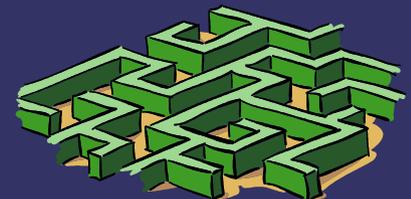
Le « *big data* » qu'est ce que c'est ?

- ⇒ Le *big data*, littéralement **grosses données** désigne l'ensemble de données que nous générons avec nos machines, nos interfaces, nos comportements... et qui sont tellement volumineuses qu'il est impossible de les traiter, les manipuler avec les outils classiques.
- ⇒ Le Big Data est tout aussi déterminant pour un pays que l'énergie ou le commerce car il peut répondre le plus efficacement aux questions de productivité et de marché .
- ⇒ Dans l'industrie, la vente, la santé ..., les outils du Big Data ont un impact immédiat en proposant des prédictions et des analyses pertinentes.



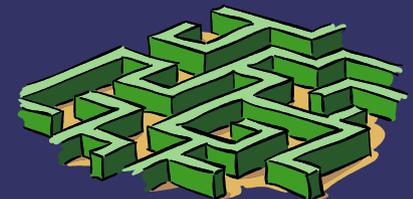
Quelques repères

- ➔ Facebook, c'est 1,3 milliard d'utilisateurs, Twitter 650 millions
- ➔ En 2016, les serveurs Facebook écoulaient 500 terra-octets de données par jour
- ➔ Les appareils connectés explosent : on compte aujourd'hui, sur Terre, une moyenne de 2,5 appareils connectés par personne
- ➔ Un Boeing 747 génère 1/2 terra-octet de données par vol



L'impact des objets connectés

- ➔ La notion d'objets connectés recouvre un panel très large qui touche tous les secteurs de l'économie
 - La santé – la maison – l'industrie – l'éducation – les loisirs – la sécurité ...
 - On peut passer du simple gadget à des outils essentiels
- ➔ Selon certaines études on estime à plus de **30 milliards le nombre d'objets connectés à l'horizon 2020 (80 milliards de « choses » connectées selon l'IDATE)**. Pour un marché estimé à **1,9 milliards de dollars**
- ➔ Le seul marché des objets connectés dédiés à la santé et à la maison, est estimé à 500M€, en France en 2016.



Différentes catégories d'objets connectés

- ⇒ On peut classer les objets connectés suivant leur finalité
 - **Les capteurs et actionneurs**
 - Dans cette catégorie on trouve de nombreux usages industriels associés à de nouveaux modèles économiques (aviation, automobile, réseaux routiers, météo, télé-relève ...)
 - Mais aussi tous les objets qui ont pour objectif de collecter des données dans des domaines aussi divers que la santé, le sport ou la sécurité (bracelets, puces détecteurs, vêtements, frigos, etc...)
 - Pour une large part des capteurs, il y a dialogue entre objets (M2M)
 - **Les « interfaces » Homme – machine**
 - Les smart phones, les ordinateurs, les téléviseurs, les outils de réalité augmentée, les cartes à puces ...



Capteurs et nouveaux modèles économiques

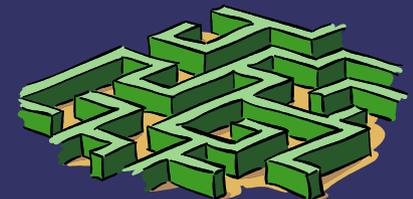
- ➔ Lors de la disparition du Boeing de la Malaysia Air Line, une autre information a été révélée : **les moteurs étaient en permanence connectés via des capteurs** et sous « surveillance » du constructeur américain Rolls-Royce.
- ➔ Les moteurs d'avions sont des technologies très sophistiquées et coûtent très chers. Ainsi, les compagnies aériennes préfèrent les louer au temps de fonctionnement.
- ➔ **Le challenge n'est plus de vendre mais de louer et de faire fonctionner des moteurs le plus longtemps possible sans panne et avec des temps d'indisponibilité les plus courts possible.**
- ➔ **Et dans la mesure où l'on est capable d'avoir des capteurs-actionneurs performants rien ne s'oppose à ce que ces modèles gagnent de nombreux secteurs de l'industrie (automobile, transport, santé ...)**



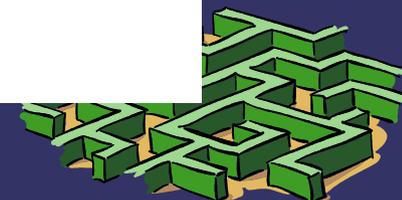
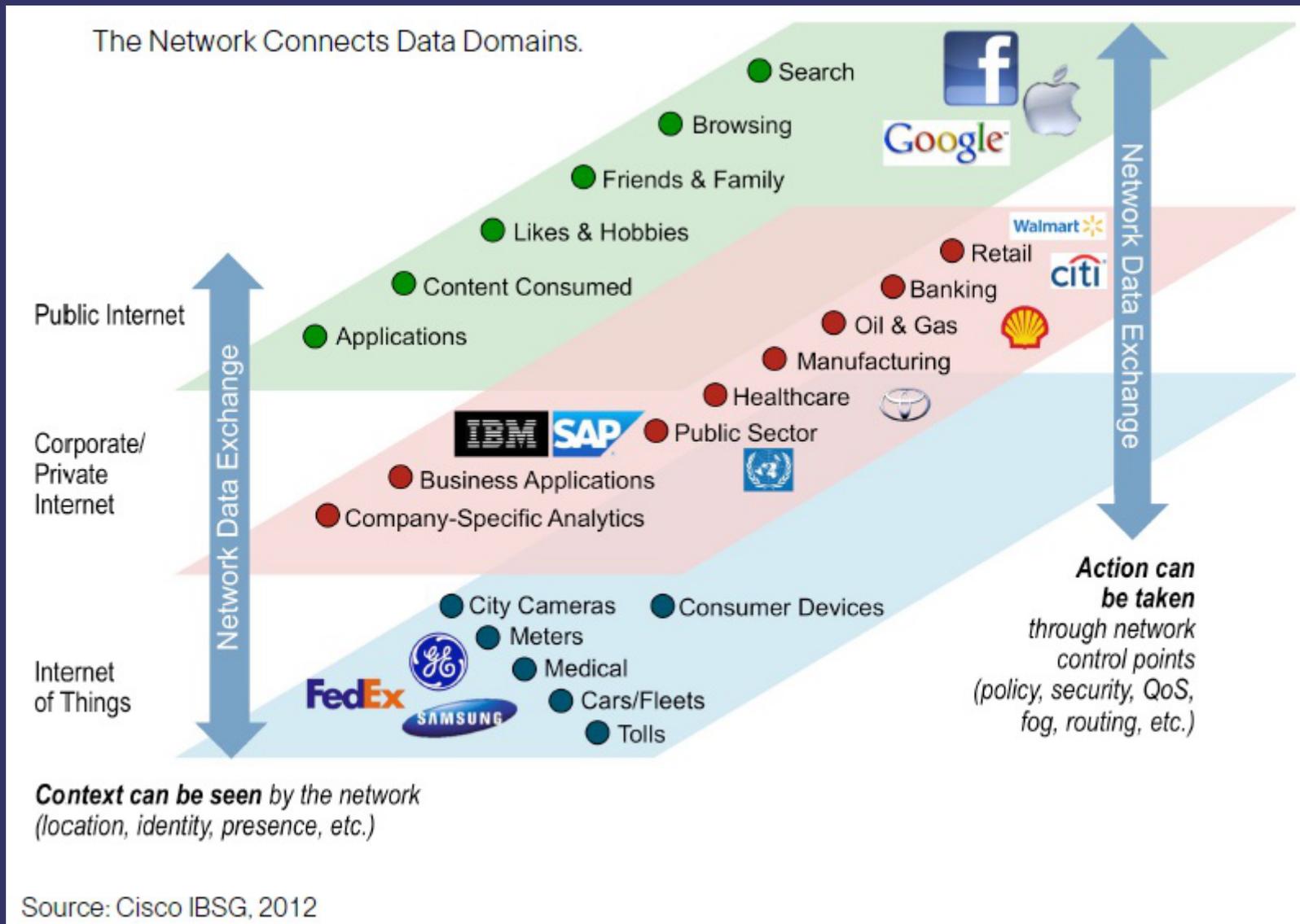
De nouvelles approches économiques

- ⇒ Dès lors que les objets ou les choses connectées remontent des informations, elles constituent de nouveaux acteurs économiques

- ⇒ Les principaux impacts sont :
 - Le marketing
 - L'optimisation des processus de production (industrie, santé, laboratoires, information...)
 - L'optimisation de la maintenance et de l'assistance
 - L'optimisation des ressources énergétiques (notamment par les réseaux intelligents)
 - Télé-relève et services associés
 - Connaissance de notre environnement (météo – pollution – prévention des risques ...)

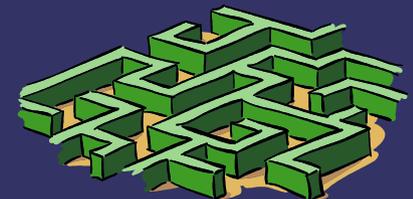


Une grande partie des données sont générées par nos comportements



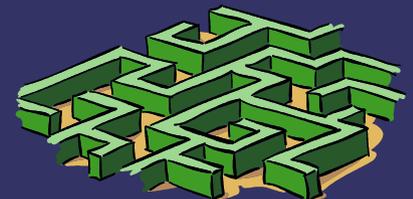
Les « interfaces » homme-machine

- ⇒ La notion d'interface Homme-machine recouvre un champ très large, dès lors que nous utilisons un équipement pour se connecter ou transférer des données, de façon consciente ou inconsciente
- ⇒ Avec ces interfaces, plus que de nouveaux modèles économiques, c'est la relation client qui est fortement impactée et de nouveaux comportements qui sont générés.
- ⇒ Maîtriser le dernier capteur, la dernière interface c'est capter la relation client et constitue un enjeu stratégique majeur
- ⇒ Mais toute action, toute utilisation, génère des données qui au delà de l'acte créent de la valeur (à titre individuel ou collectif)
- ⇒ L'exploitation de ces données pose alors des problèmes sociaux



Nouveaux comportements et usages :

- ⇒ Le marketing ciblé (assurance, vente, publicité etc.)
- ⇒ La santé
- ⇒ Le tourisme
- ⇒ Le coatching sportif (*fitness ou données médicales*)
- ⇒ La sécurité des personnes
- ⇒ ...



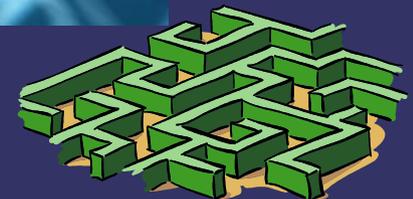
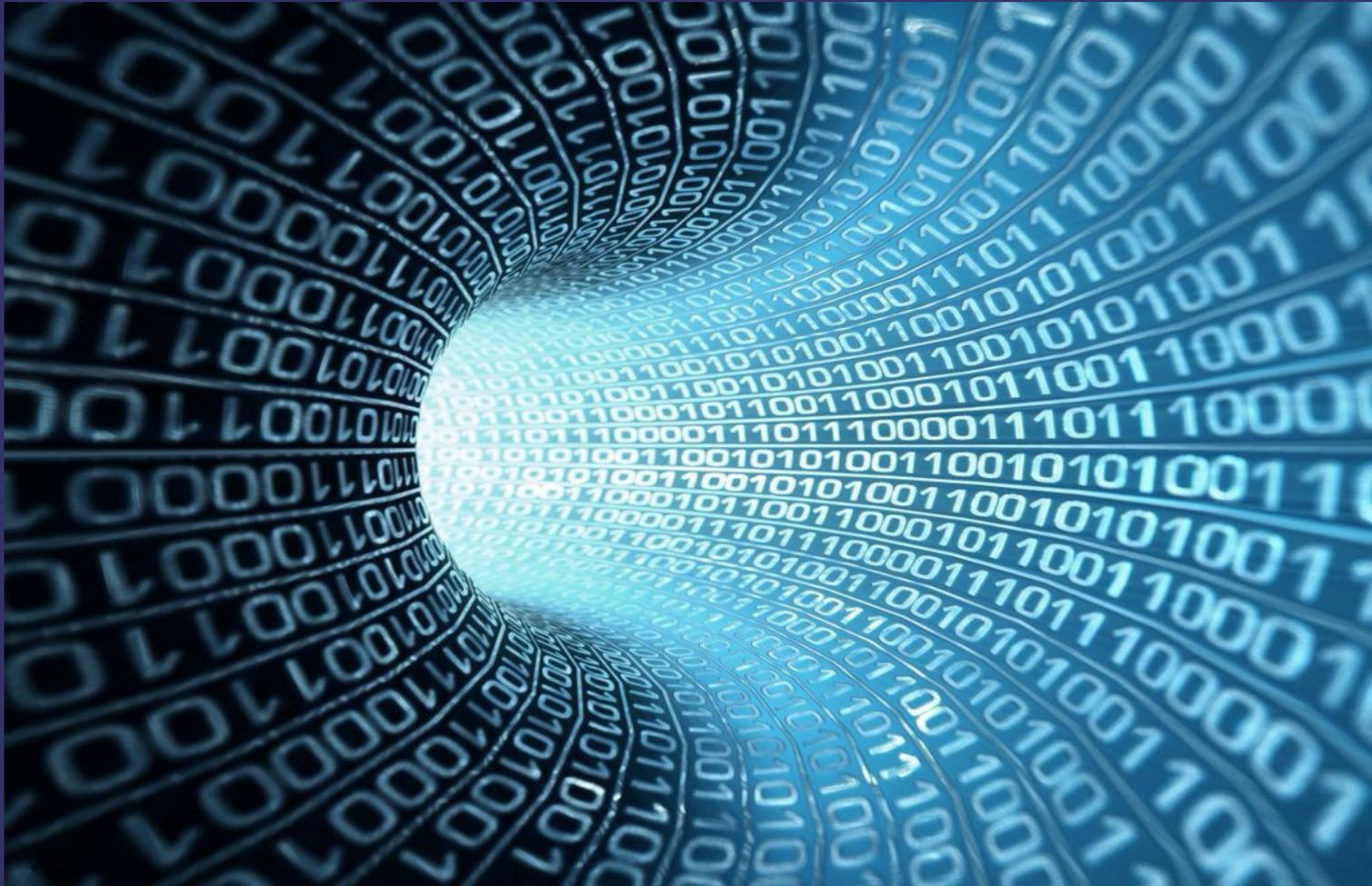
La révolution des objets connectés ...

- ⇒ La capacité des réseaux, les énormes progrès réalisés au niveau des micro-processeurs et du stockage ouvrent la voie des objets connectés.
- ⇒ Il y aura donc explosion des objets connectés dans tous les domaines économiques et notre quotidien.
- ⇒ Les objets et capteurs relevant de l'interface homme-machine sont poussés par l'innovation et de très nombreuses start-up.
- ⇒ Mais la normalisation des interfaces et des standards d'échange constituent une bataille économique à elle seule.
- ⇒ Par ailleurs, il y a un risque réel de « bulle ». Il faut donc s'attendre à une rationalisation puis à une croissance mature de ce marché.
- ⇒ C'est une véritable révolution industrielle qui est en marche profondément liée à la capacité de stockage et de traitement des données produites.



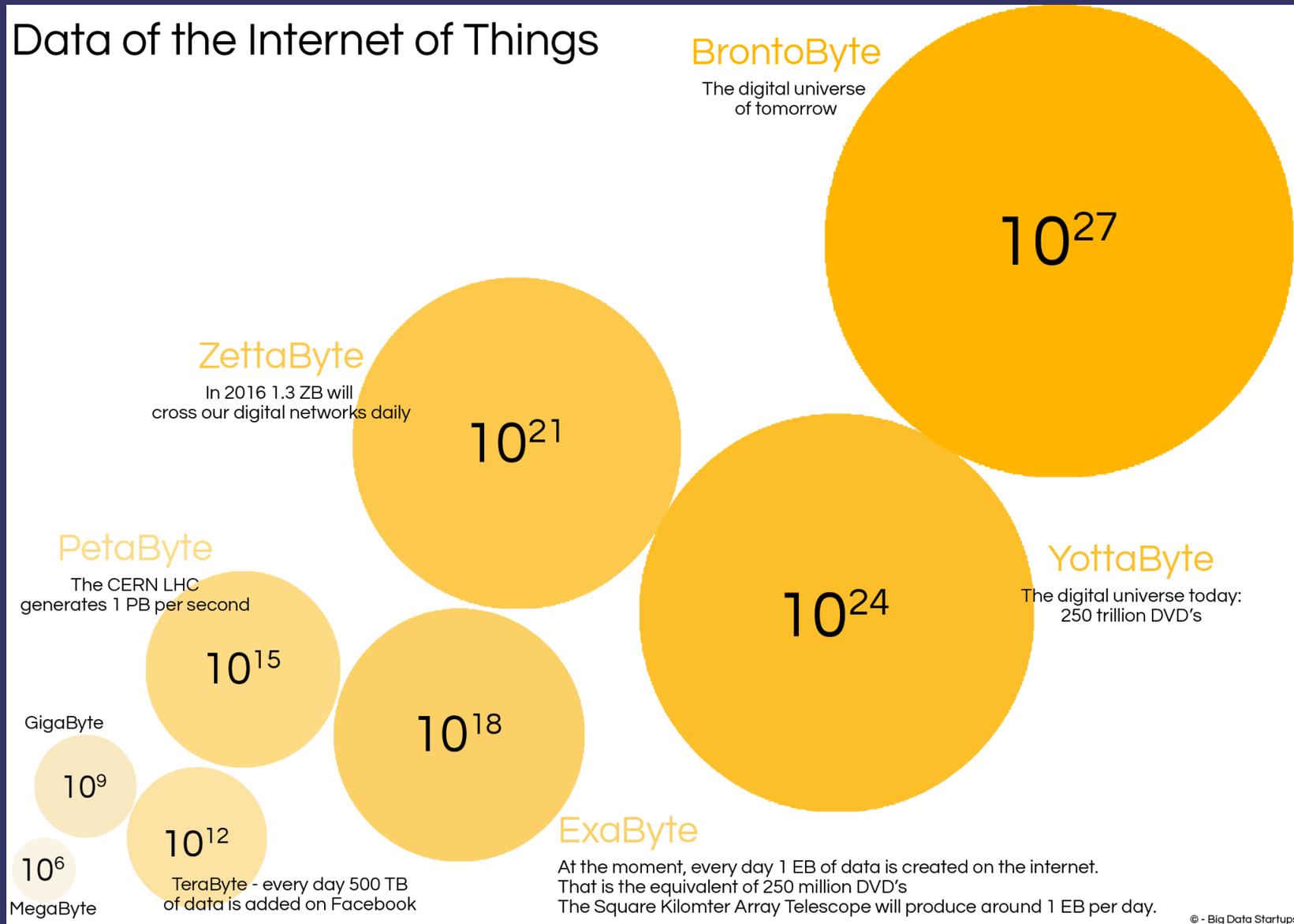
2

Le traitement et l'exploitation des données



Quelques chiffres

Data of the Internet of Things



© - Big Data Startups



Quelques chiffres ...

www.wipro.com

BIG DATA

Big Data is data that is too large, complex and dynamic for any conventional data tools to capture, store, manage and analyze.

The right use of Big Data allows analysts to spot trends and gives niche insights that help create value and innovation much faster than conventional methods.

The "three V's", i.e the Volume, Variety and Velocity of the data coming in is what creates the challenge.

VOLUME



Amount of Big Data stored across the world (in petabytes)

VARIETY



PEOPLE TO PEOPLE

NETIZENS, VIRTUAL COMMUNITIES, SOCIAL NETWORKS, WEB LOGS...



PEOPLE TO MACHINE

ARCHIVES, MEDICAL DEVICES, DIGITAL TV, E-COMMERCE, SMART CARDS, BANK CARDS, COMPUTERS, MOBILES...



MACHINE TO MACHINE

SENSORS, GPS DEVICES, BAR CODE SCANNERS, SURVEILLANCE CAMERAS, SCIENTIFIC RESEARCH...



2.9 MILLION EMAILS SENT EVERY SECOND



20 HOURS OF VIDEO UPLOADED EVERY MIN

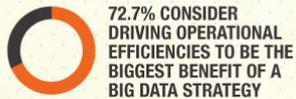
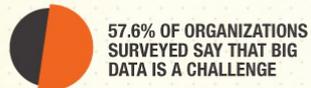
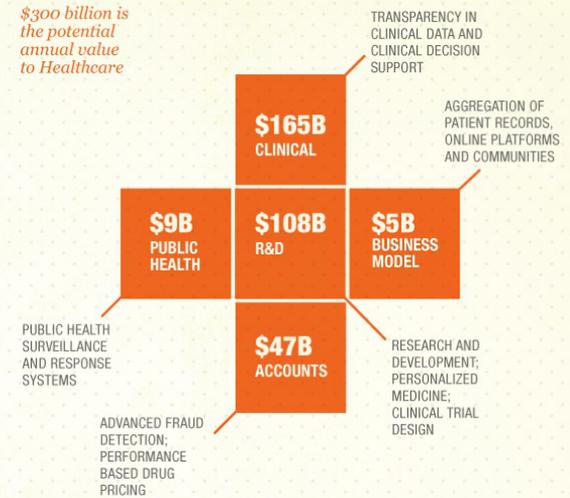


50 MILLION TWEETS PER DAY

VELOCITY

CASE STUDY - Healthcare

\$300 billion is the potential annual value to Healthcare



VALUE



40% PROJECTED GROWTH IN GLOBAL DATA CREATED PER YEAR



5% PROJECTED GROWTH IN GLOBAL IT SPENDING PER YEAR

The estimated size of the digital universe in 2011 was 1.8 zettabytes. It is predicted that between 2009 and 2020, this will grow 44 fold to 35 zettabytes per year. A well defined data management strategy is essential to successfully utilize Big Data.

Sources: Reaping the Rewards of Big Data - Wipro Report; Big Data: The Next Frontier for Innovation, Competition and Productivity - McKinsey Global Institute Report; numScore, Radical Group; Measuring the Business Impacts of Effective Data - study by University of Texas, Austin; US Department of Labour.

DO BUSINESS BETTER

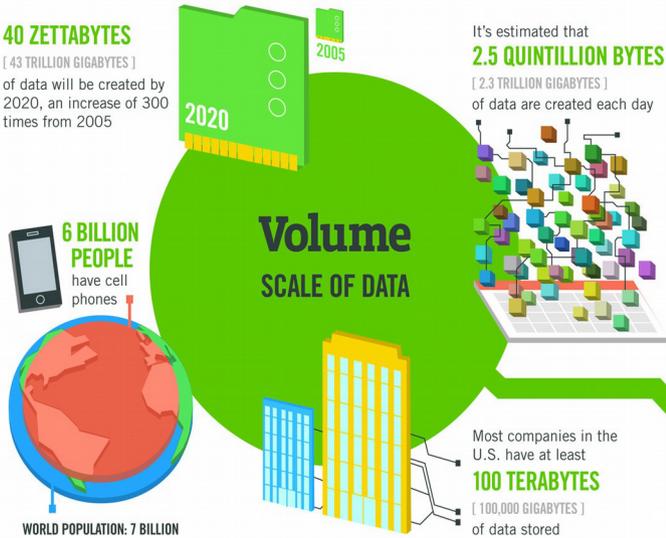
NYSE:WIT | OVER 130,000 EMPLOYEES | 54 COUNTRIES | CONSULTING | SYSTEM INTEGRATION | OUTSOURCING



Les enjeux

40 ZETTABYTES

[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES
[161 BILLION GIGABYTES]



30 BILLION PIECES OF CONTENT
are shared on Facebook every month



By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month



400 MILLION TWEETS
are sent per day by about 200 million monthly active users



Variety DIFFERENT FORMS OF DATA

The New York Stock Exchange captures
1 TB OF TRADE INFORMATION
during each trading session



Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

— almost 2.5 connections per person on earth



Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure

1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions



Poor data quality costs the US economy around
\$3.1 TRILLION A YEAR



in one survey were unsure of how much of their data was inaccurate

Veracity UNCERTAINTY OF DATA

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, QAS



Une révolution industrielle

Les organisations utilisant activement le Big Data bénéficient d'un taux de croissance du chiffre d'affaires 50 % plus élevé que celles qui ne l'utilisent pas

LES COÛTS REPRÉSENTENT LE PRINCIPAL OBSTACLE POUR LES ORGANISATIONS N'UTILISANT PAS ENCORE LE BIG DATA, MAIS AUSSI POUR CELLES QUI L'UTILISENT ACTUELLEMENT

PRINCIPAUX OBSTACLES POUR CELLES QUI N'UTILISENT PAS ENCORE LE BIG DATA

20 %

Ne savent pas si les avantages justifient le coût

18 %

Coût de l'infrastructure informatique

17 %

Coût de la sous-traitance des opérations ou des analyses

PRINCIPAUX OBSTACLES POUR CELLES QUI UTILISENT ACTUELLEMENT LE BIG DATA

29 %

Coût de l'infrastructure informatique

26 %

Coût de la sous-traitance des opérations ou des analyses

PRINCIPAUX RÉSULTATS DE L'UTILISATION DU BIG DATA



Meilleur ciblage des efforts marketing

41 %



Optimisation des dépenses en publicité

37 %



Optimisation du marketing dans les médias sociaux

37 %

44 % des organisations dans le monde ne savent toujours pas comment aborder le Big Data, mais les choses évoluent, particulièrement en Amérique du Nord :

De plus en plus d'organisations d'Amérique du Nord estiment disposer de données Big Data pouvant être analysées

54 % en 2014 → 73 % en 2015



Les organisations d'Amérique du Nord estiment également tirer un meilleur parti de leurs données, en ayant atteint

64 % → 69 %

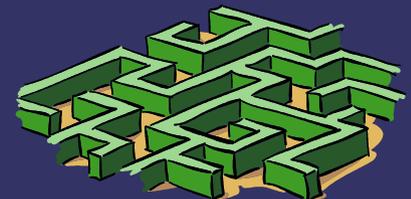
du potentiel des données en 2014 du potentiel des données en 2015

➔ D'après une étude DELL-EMC, les entreprises européennes ont peu pris conscience de l'intérêt de l'exploitation des données contrairement à l'Amérique du Nord ou aux Etats Unis.



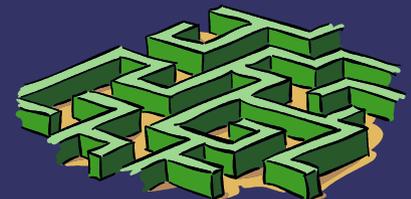
Des bouleversements dans les entreprises

- ⇒ On passe d'un management basé sur l'expérience et l'intuition à un management plus rationnel
- ⇒ Sur les décisions stratégiques
- ⇒ Sur la relation client
- ⇒ Sur la connaissance client et le marketing
- ⇒ Sur la sécurité et la protection des données



Des bouleversements sociétaux

- ⇒ Parce que nous produisons sans le vouloir des volumes colossaux de données
- ⇒ Données dont on n'est pas toujours propriétaires. Une récente décision de **Donald Trump qualifie les données produites par les particuliers comme des données publiques.**
- ⇒ Ces données ont une valeur marchande importante
- ⇒ Avec la généralisation du cloud, **la protection des données est devenue un enjeu important**
- ⇒ **La protection de la vie privée** est un sujet majeur de nos démocraties.



Focus sur la France

- Dès 2013, le rapport d'Anne Lauvergeon « Innovation 2030 » identifie clairement cet enjeu stratégique
- Selon une étude EMC auprès des entreprises françaises, le Big Data facilite,
 - ⌘ la prise de décision pour 74 %
 - ⌘ l'ascension des leaders du marché pour 47 %
 - ⌘ des avantages compétitifs pour 23 %
 - ⌘ La cyber attaque pour 40 %

Mais ...

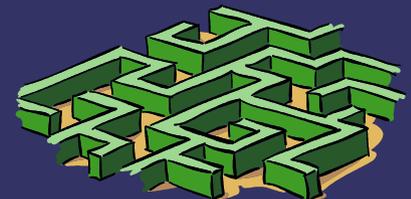
- ⌘ **61 % considèrent que le budget est limitatif et 41 % ne prévoient aucun budget.**
- ⌘ Les principaux freins identifiés par les entreprises :
 - 1- le ROI
 - 2- la pertinence
 - 3- la culture



Quelques exemples

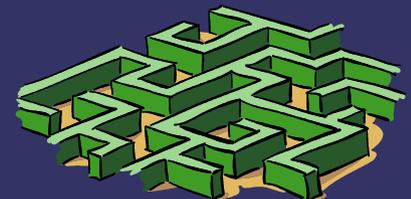
Le domaine de la santé :

- ⇒ D'après Orange Healthcare, les données de santé devraient être **multipliées par 50 d'ici 2020**
- ⇒ **L'exploitation intelligente des données massives** permet de réaliser très rapidement des analyses comparatives sur d'immenses quantités d'informations, ce que l'être humain seul ne peut pas faire.
- ⇒ **Ainsi émerge une nouvelle médecine** dite « 4P », c'est-à-dire Prédictive, Préventive, Personnalisée et Participative.



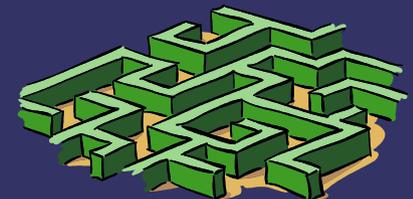
Quelques exemples

- ⇒ **Netflix**, par l'analyse de plusieurs années de données d'utilisateurs s'est totalement transformée pour devenir un créateur de contenus à succès et une machine à cash.
- ⇒ **Dans l'industrie automobile**, la voiture connectée remonte des données majeures pour fiabiliser les modèles, mieux cibler les usages, optimiser les chaînes de production en évitant les retours usines coûteux et offrir de nouveaux services aux clients.

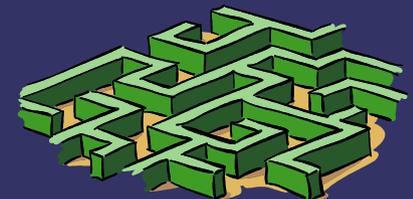
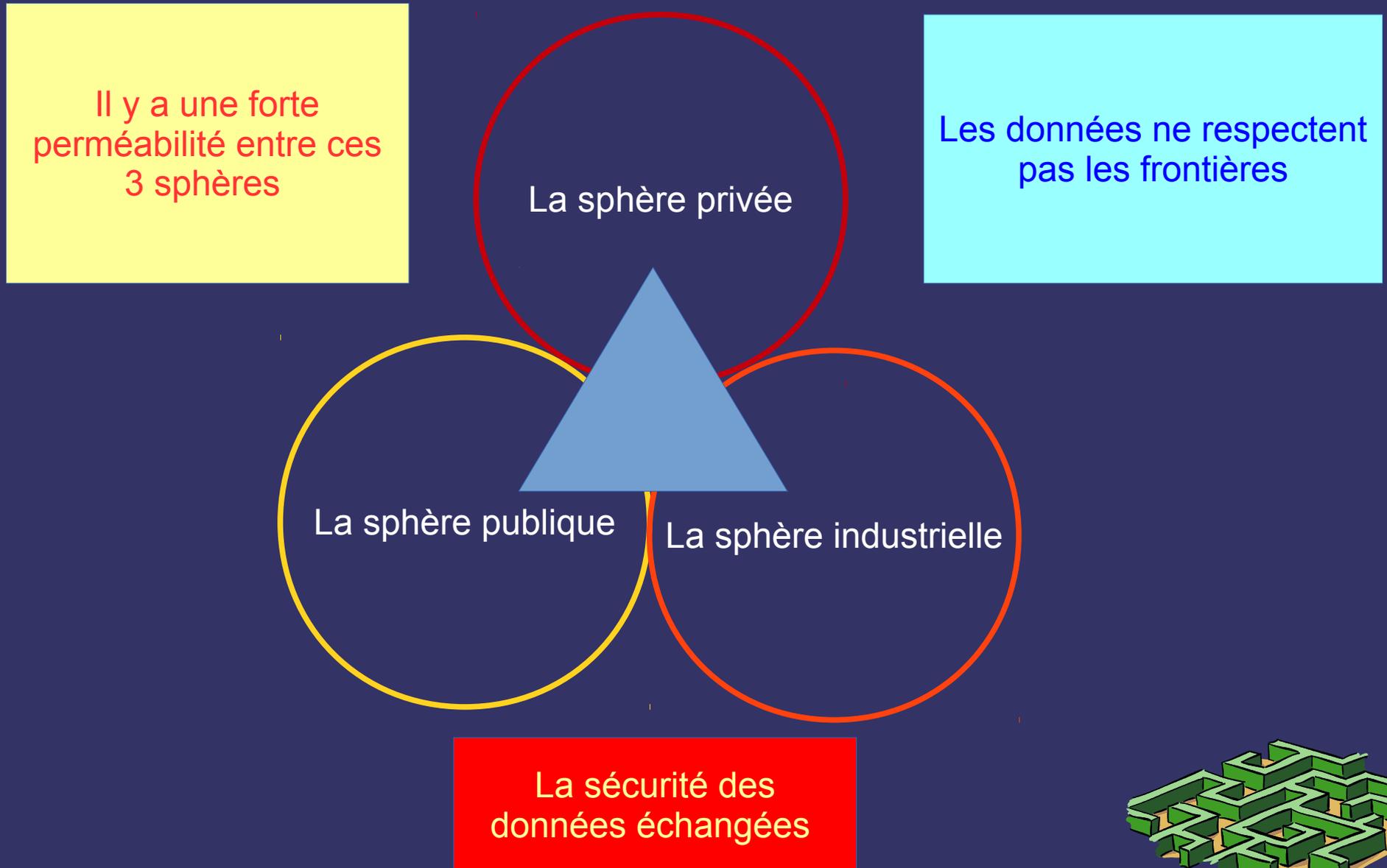


3

Les opportunités, les risques d'un monde ouvert



Trois grandes sphères de gisements de données



Le big data, des enjeux de souveraineté

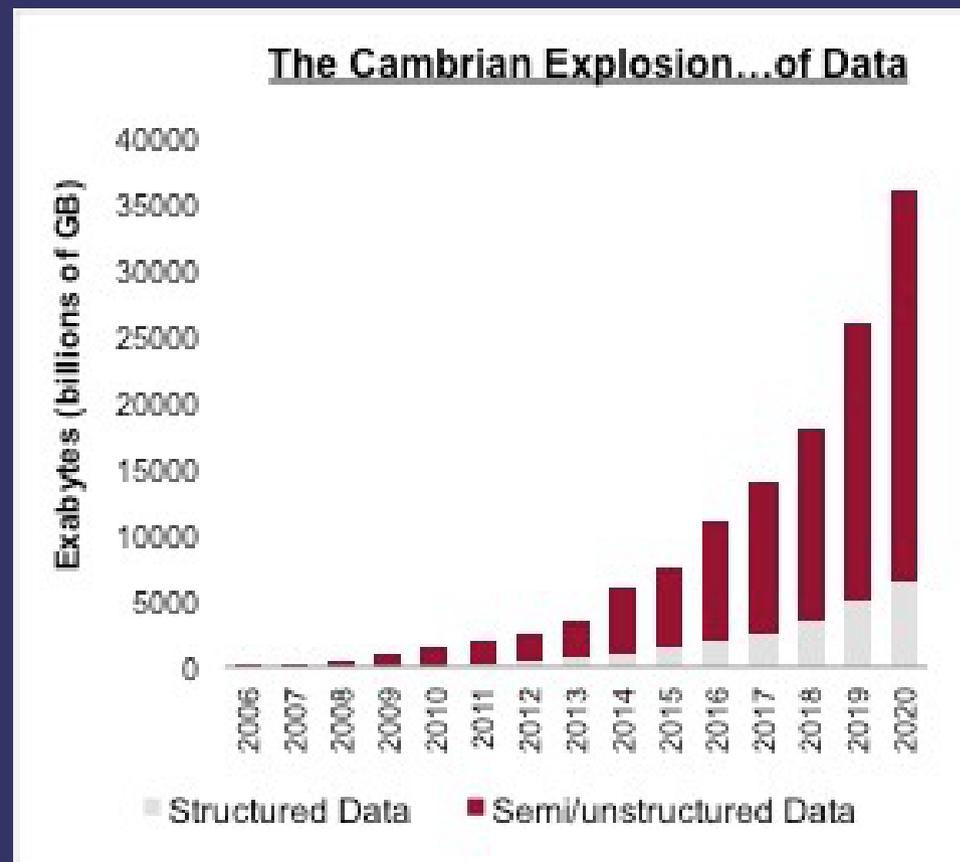
➤ Avec des prévisions « astronomiques » de 35 zettabytes de données produites en 2020 et des capacités de stockage de 15 zettabytes, **deux grands enjeux se dessinent pour une indépendance économique et stratégique:**

➤ **La capacité à stocker** les données produites et exploitables

➤ **La capacité à traiter et à utiliser** ces immenses gisements de données

✎ De gros moyens de recherche sont et devront être investis

✎ C'est presque une filière industrielle à part entière



Les opportunités

- ⇒ Elles sont réelles et nombreuses
 - L' économie,
 - En termes d'emplois
 - De nouveaux modèles économiques
 - D'une meilleure compétitivité des entreprises
 - Une plus grande pertinence des choix stratégiques
 - La connaissance de notre environnement
 - Météo
 - Pollutions
 - Connaissance des phénomènes extrêmes
 - Des progrès sur la connaissance et le comportement des populations
 - Santé – épidémies ..
 - Comportement – mobilité ...
 - Le développement de nouveaux services personnalisés (santé, sécurité, gestion des données personnelles ...)



Les risques

- ⇒ Le risque majeur est la perte de maîtrise de nos données au profit de géants privés.
 - Avec à la clé
 - Une perte de souveraineté
 - Une dépendance économique
- ⇒ Une concurrence économique au profits de géants mondiaux
 - Avec des risques réels pour les PME
- ⇒ Une prédominance de la finance dans les grands dossiers régaliens
- ⇒ Un cadre réglementaire et législatif (national et Européen) inadapté aux enjeux
- ⇒ Le risque d'intrusion dans la sphère privée
 - Avec des conséquences innombrables dans le domaine de la santé, l'éducation, la culture,
 - Et finalement la liberté et la démocratie

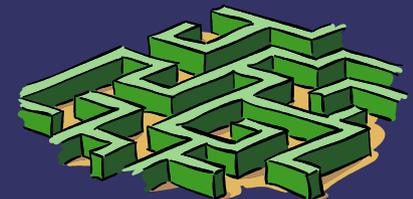


En conclusion

- ➔ Les données produites passent par les réseaux
 - **Le déploiement du très haut débit est essentiel**, cependant, les investissements sont très importants (fibre optique – 5G – réseaux de collecte...) et le cadre réglementaire pas suffisamment incitatif
- ➔ **Maîtriser la relation client** c'est favoriser l'émergence d'une puissante innovation déjà existante et qu'il faut favoriser
- ➔ **Le stockage de nos propres données et leur traitement** sont un enjeu stratégique majeur pour notre pays (pour l'Europe?) et supposent de très lourds investissements et des programmes de recherche qu'il convient de renforcer
- ➔ Nos entreprises, bien que conscientes, n'ont pas encore pris toute la mesure des risques concurrentiels d'un monde totalement ouvert où **celui qui maîtrise les données maîtrise le marché.**



Le pétrole de demain ?



Merci de votre attention

