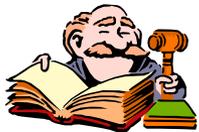




## Journée Annuelle - 8 décembre 2015

« *Définition du Big Data* »

Jean Pierre MALLE



# m8

## jean-pierre malle

- datascientist, expert data-intelligence
- modèles et algorithmes brevetés :
  - en analyse contextuelle
  - en analyse comportementale
  - en analyse cognitive
  - en analyse situationnelle
- conseil scientifique au board de 5 pme innovantes en data-intelligence
- formations et conférences :
  - écoles, universités, entreprises, meet'up
  - ionix : « data science et analyse situationnelle dans les coulisses du big data »

données  
humaines et  
sociales





## sommaire

- le big data
- les métiers du big data
- le marché de la data
- la personnalisation de l'information
- les machines apprenantes
- pour en savoir plus





PRISE DE REPERES

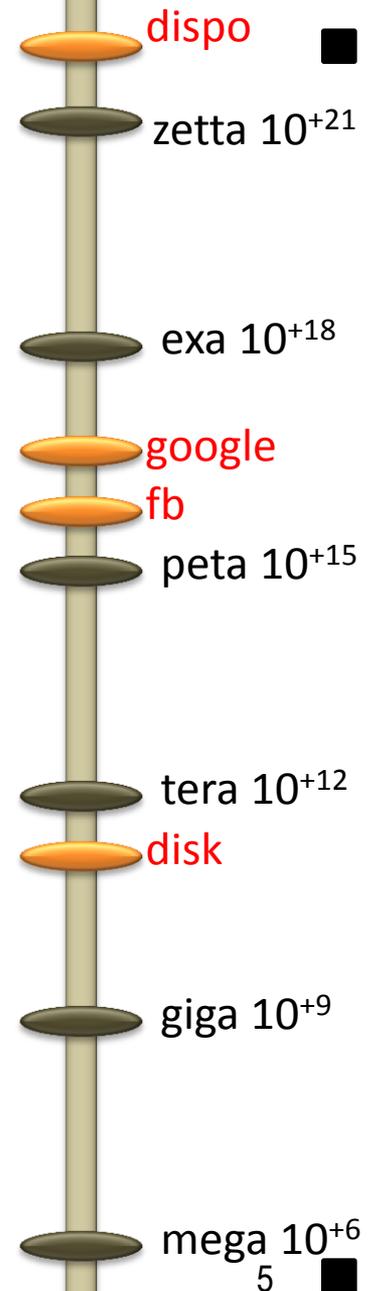
# LE BIG DATA



# m8

## ou commence le big data ?

- on estime à fin 2015 que :
  - le nombre de données digitales **disponibles** à 7 zb, cela représente une pile de papier de 70 fois la distance terre-soleil
  - la **production** de données affiche une croissance de 40% par an
  - google a traité 250 petabytes de données par jour (a ce rythme il faut 80 ans pour traiter le stock **disponible** aujourd'hui).
  - facebook brasse et analyse plus de 150 petabytes de données chaque jour.

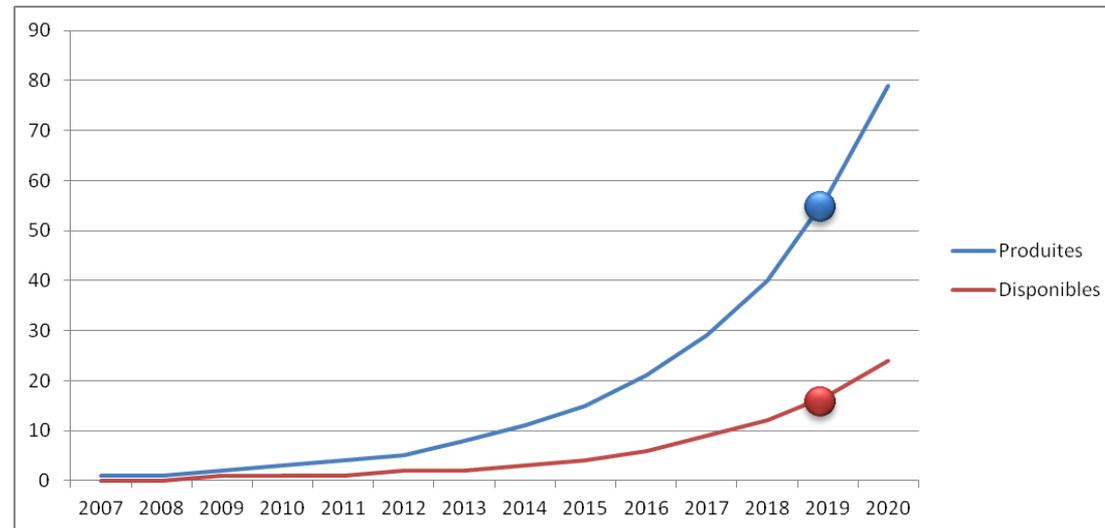


Extrait : IonisX : MOOC Datascience dans les coulisses du Big Data

par JP Malle

## changement de paradigme

- chaque année on produit presque autant de données que depuis le début de l'humanité (75%)
- 90% des données disponibles ont moins de 2 ans
- c'est un système ouvert et en expansion



Extrait : IonisX : MOOC Datascience dans les coulisses du Big Data par JP Malle



# m8

## les 3 v

- volume

- on ne peut pas tout stocker
- on ne peut pas tout scanner

incomplétude

- vélocité

- on ne peut pas tout suivre
- les données sont relatives

incohérence

- variété

- les données sont non structurées
- les codes et les sens évoluent

ambiguïtés



Source : G9+ : Big Data, l'accélérateur d'innovation

## les v en plus

- véracité
  - on ne peut pas tout vérifier
  - les données ne sont pas tracées
- visualisation
  - on ne peut pas tout voir
  - les représentations sont réductrices
- valorisation
  - la valeur dépend des usages
  - la valeur dépend des sources

incertitudes

parcellarité

estimation



Source : G9+ : Big Data, l'accélérateur d'innovation

m8

une définition



Le big data

*c'est travailler avec des  
données incertaines*



*hier*

*aujourd'hui*

● conserver plus de données pour plus de précision

● oublier plus de données pour plus de possibilités

● éliminer les cas singuliers pour se concentrer sur les plus fréquents

● abandonner les cas fréquents pour se concentrer sur les différentiateurs

● modéliser et normaliser les données

● rechercher les singularités et les inconnues

● traiter exhaustivement les données

● exploiter les données essentielles



Extrait : Paris Tech Review : La triple rupture des big data par JP Malle

- big data
  - données volumiques
- smart data
  - données internes
- small data
  - données ciblées
- open data
  - données partagées
- données non structurées
- données non garanties
- données évolutives
- données interprétables
- données parfois erronées, incomplètes, ambiguës, ...



Source : Guide du Big Data 2014/2015



PRISES DE REPERES

# LES MÉTIERS DU BIG DATA



## le datascientist

- connaît :
  - le métier pour comprendre les données
  - les techniques et outils d'analyse
  - les sources de données
- n'est pas :
  - un dataminer
  - un statisticien
  - un sémanticien
- il doit maîtriser des notions telles que :
  - abstraction
  - connexité
  - essentialité

des modèles  
qui  
s'expliquent

des modèles  
qui agissent



Extrait : seekube.com : Etre datascientist c'est quoi ? par JP Malle

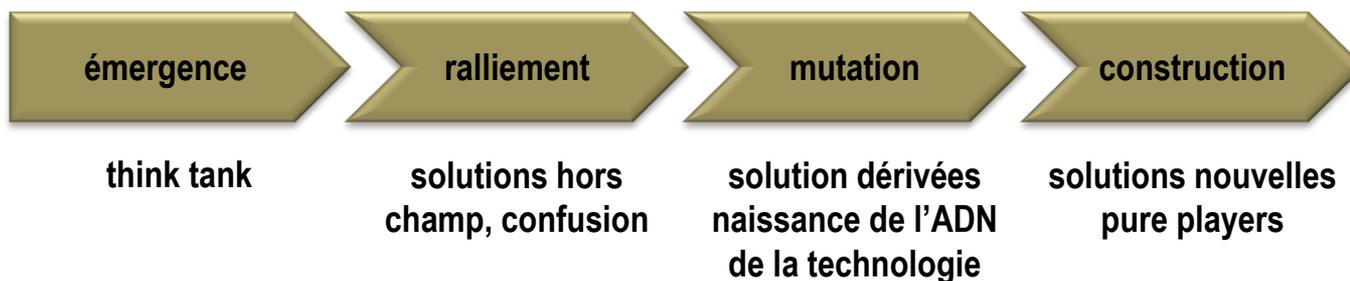
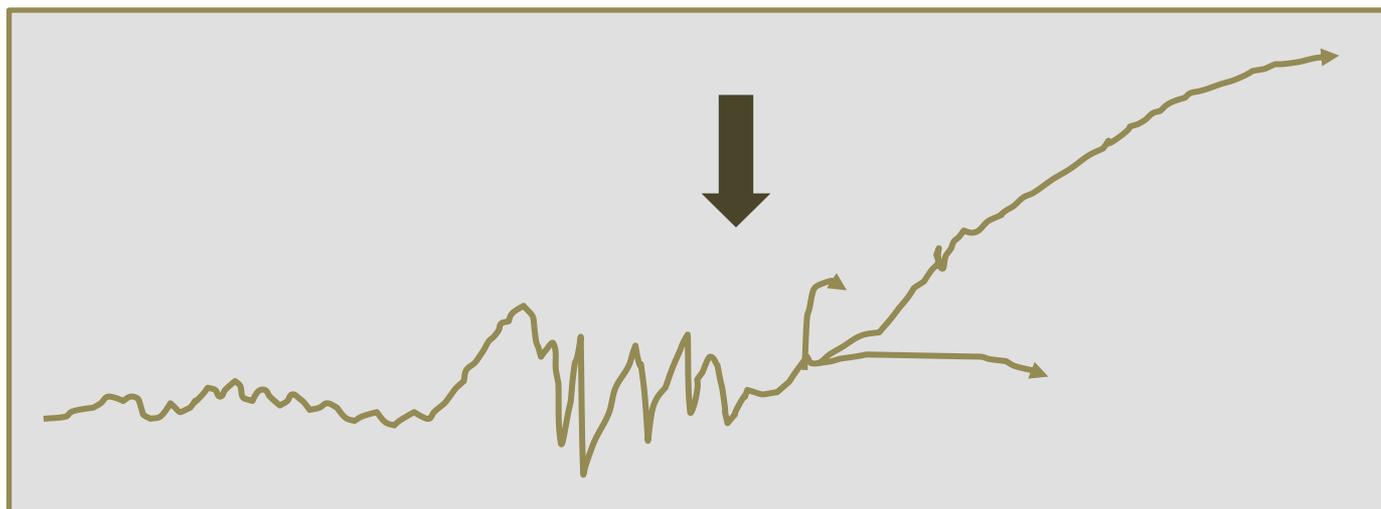


## le CDO

- au quotidien :
  - il valorise les données
  - il conçoit des data-services (usages des données)
  - il veille au respect de la vie privée
- positionnement :
  - idéalement rattaché à la DG
  - chaînon manquant métiers-informatique
  - profil visionnaire
- il doit :
  - comprendre les métiers
  - agir avec célérité (données périssables)
  - disposer d'un portefeuille de sources de données



Extrait : Silicon.fr : Le CDO est le chaînon manquant entre directions métiers et DSI par JP Malle



Extrait : Paris Tech Review : La triple rupture des big data par JP Malle



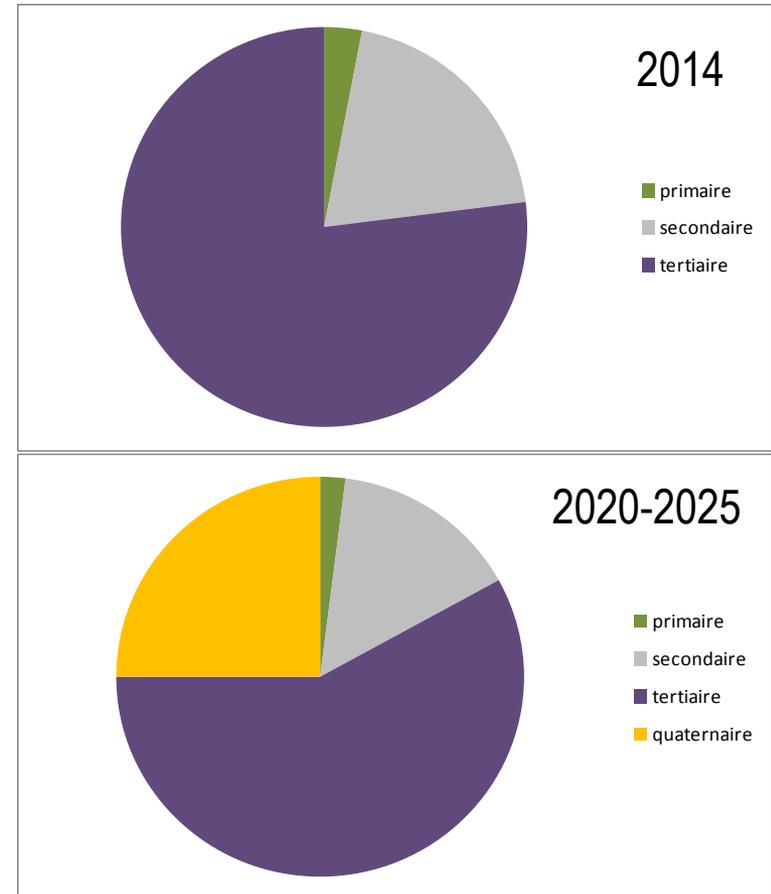
PRISES DE REPERES

# LE MARCHE DE LA DATA



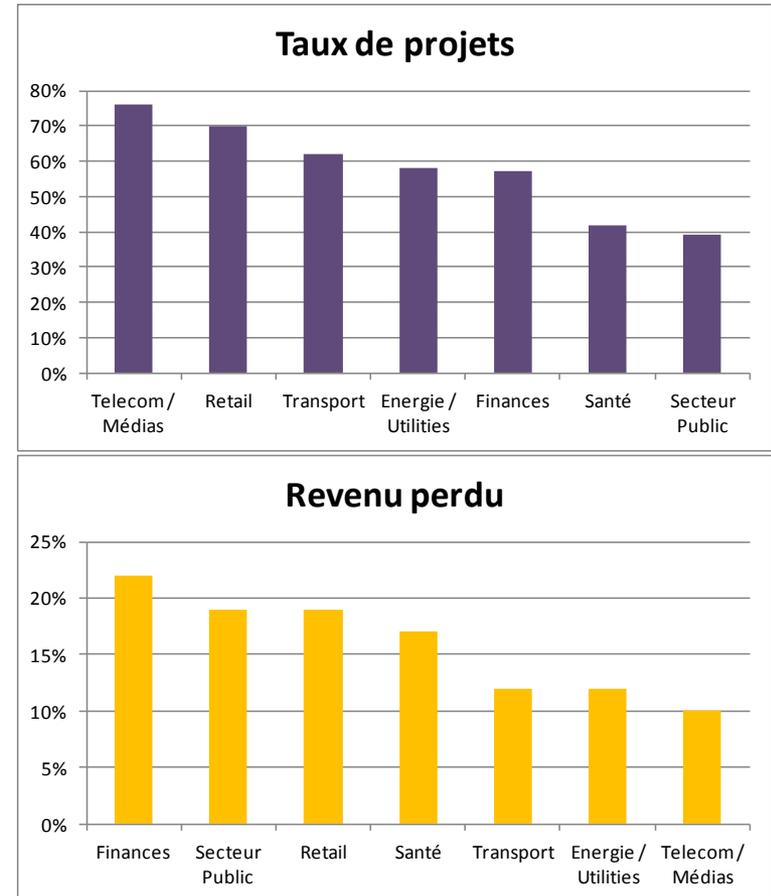
## secteur quaternaire ?

- Secteur primaire 2%
  - (agriculture, pêche, forêts, mines)
- Secteur secondaire 20%
  - (industrie)
- Secteur tertiaire 78%
  - (services)
- Secteur quaternaire
  - (données)
  - 25% du PIB à l'horizon H2020-2025



Extrait : IonisX : MOOC Datascience dans les coulisses du Big Data par JP Malle

- Taux d'entreprises du secteur ayant eut au moins un projet big data en 2014
- Revenu perdu estimé par les entreprises du secteur du fait de leur retard big-data



Source : data-business.fr à l'aide d'informations produites par SAP et Oracle



PRISES DE REPERES

# PERSONNALISATION DE L'INFORMATION





## la personnalisation

- l'heure est à la prise en compte des **individualités**
- les modèles statistiques fédérateurs laissent la place aux modèles **différentiateurs** captant les signaux faibles et **apprenant**
- l'intérêt est de prévoir comment **chacun** va raisonner par rapport à **sa** vision de la situation, **son** référentiel , **sa** rationalité, **ses** biais cognitifs...
- **mieux connaître** un individu, c'est mieux le servir, jouer sur ses sensibilités, anticiper ses réactions



Extrait : IonisX : MOOC Datascience dans les coulisses du Big Data par JP Malle



## enquête sites e-commerce

- 2/3 internautes veulent de la personnalisation
  - 67% préfèrent acheter sur un site avec offres personnalisées
  - 63% recommandent davantage un site avec offres personnalisées
  - 69% des internautes attendent des Promotions personnalisées sur les sites avec personnalisation
- parmi eux 1/3 sont inquiets pour leurs données
  - 32% des internautes ne sont pas rassurés par la personnalisation proposée



Source : livre blanc Netwave + Markess International

## Le PDG de Google prédit la fin de l'anonymat sur Internet

Le Monde.fr | 05.08.2010 à 17h15



Eric Schmidt, le directeur exécutif de Google. | AP/Paul Sakuma





PRISES DE REPERES

# MACHINES APPRENANTES

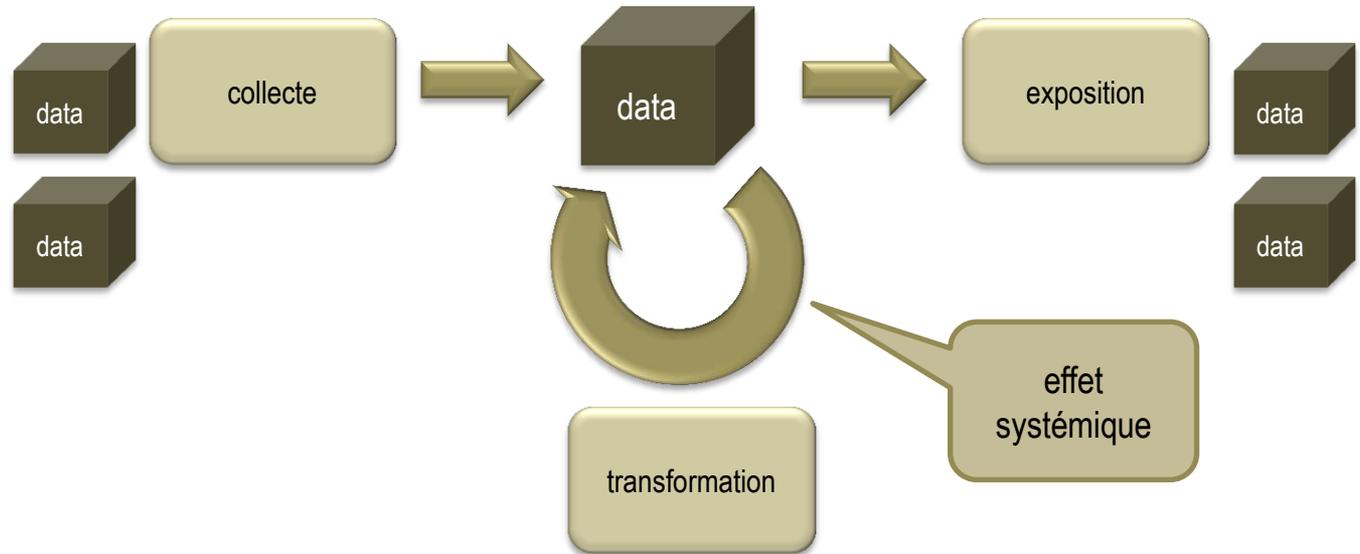


## l'homme et le big data

- l'homme est une machine bigdata
- l'homme a développé des mécanismes pour traiter de grands volumes d'information
  - en temps réel
  - en temps différé
  - en temps révélé
- la technologie se rapproche de plus en plus du modèle humain



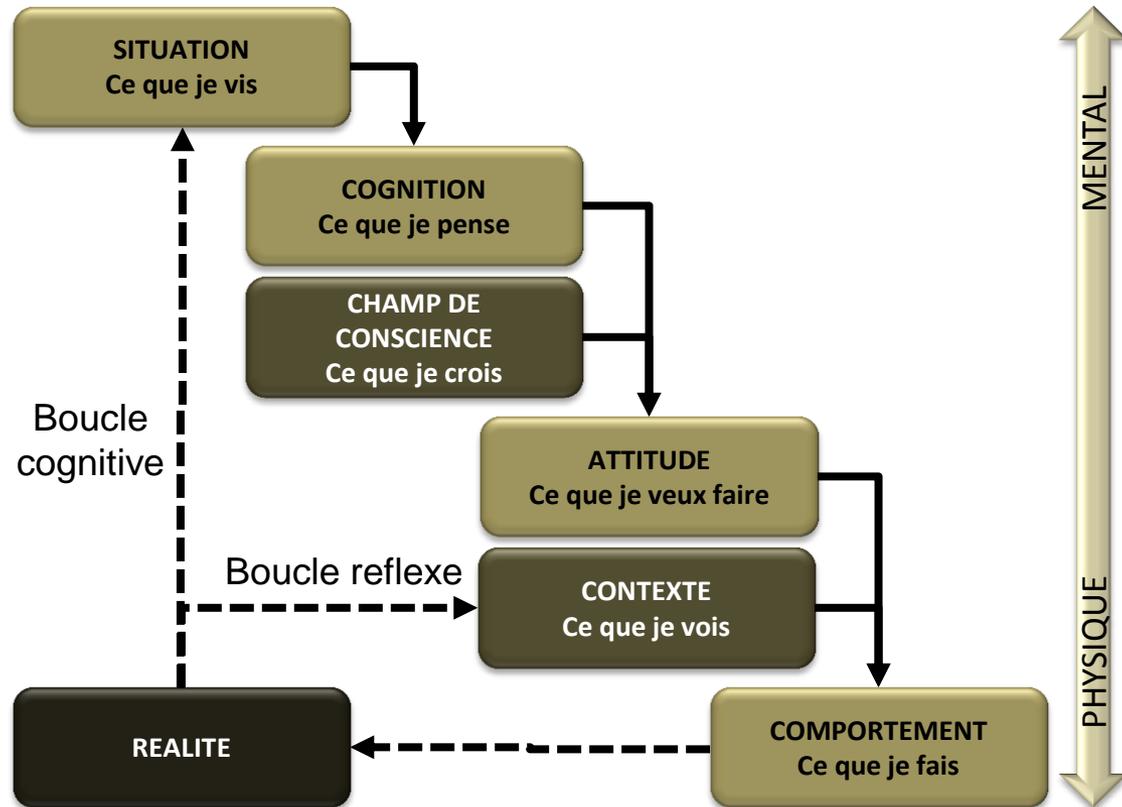
Extrait : IonisX : MOOC Datascience dans les coulisses du Big Data par JP Malle



Extrait : IonisX : MOOC Datascience dans les coulisses du Big Data par JP Malle

# m8

## exemple : machine apprenante d'analyse situationnelle



Extrait : IonisX : MOOC Datascience dans les coulisses du Big Data par JP Malle

m8



**POUR EN SAVOIR PLUS**



m8

pour en savoir plus

IONISX



Cours gratuits (MOOC)

<https://ionisx.com/>



**Datascience : dans les coulisses du Big Data**

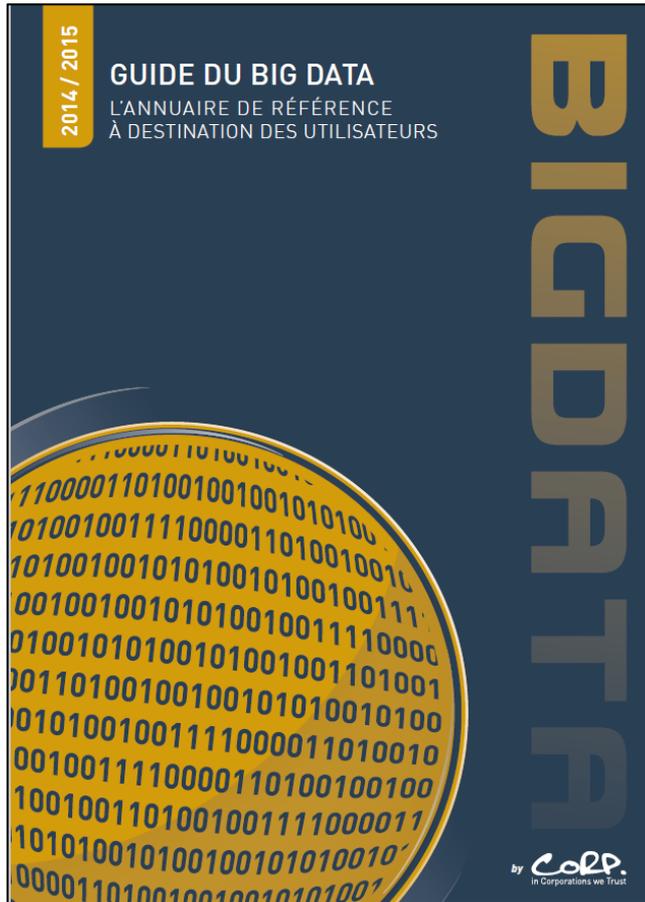
datascience, big, data

La datascience et l'analyse situationnelle : l'étude des Big Data

INSCRIPTION

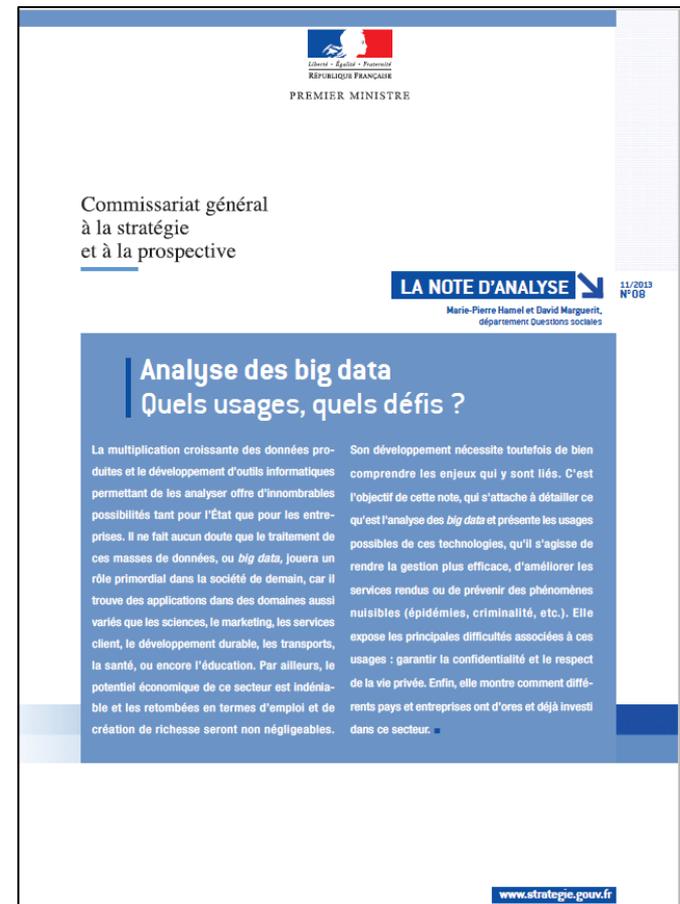
m8

pour en savoir plus



# m8

# pour en savoir plus



## Premiers Pas dans les Big Data

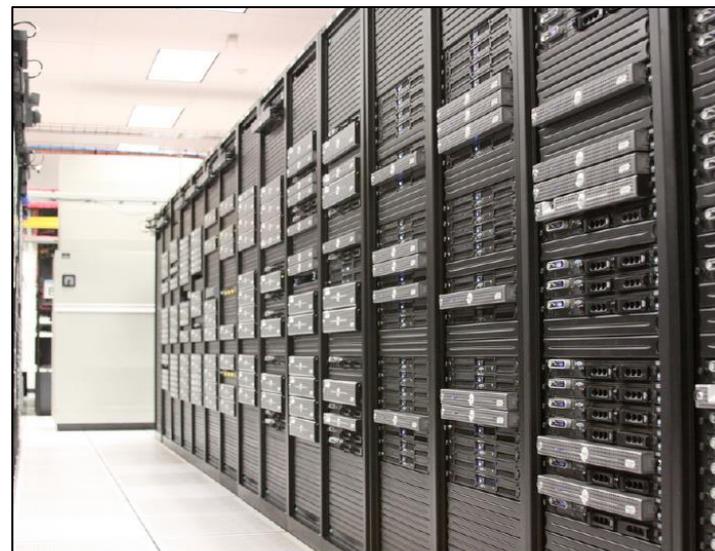
Michel Bruley

<p><b>Text Mining</b></p>	<p><b>Sentiment Analysis</b></p>	<p><b>Social Network</b></p>
<p><b>Web log &amp; Clickstream</b></p>	<p><b>MapReduce</b></p>	<p><b>Marketing Attribution</b></p>
<p><b>Social CRM</b></p>	<p><b>Churn</b></p>	<p><b>Machine Learning</b></p>
<p><b>Product Affinity</b></p>	<p><b>Next best offer</b></p>	<p><b>Data Visualization</b></p>
<p><b>GeoMarketing</b></p>	<p><b>Pricing</b></p>	<p><b>Natural Language Processing</b></p>

Michel Bruley

<http://www.decideo.fr/bruley/>

1/32



VERSION 1.0 - MARS 2012

## BIG DATA & OPEN SOURCE: UNE CONVERGENCE INÉVITABLE?

**Stefane Fermigier**



m8



 jpmalle

**MERCI**

