

18 mars 2014



Big Data : Risques et contre-mesures

Les fondamentaux pour bien démarrer

Gérôme BILLOIS

gerome.billois@solucom.fr

Twitter : @gbillois

Chadi HANTOUCHE

chadi.hantouche@solucom.fr

Twitter : @chadihantouche

solucom 
management & IT consulting

- ▶ 1. Le Big Data vous pensez que c'est flou?
- 2. Les grands risques du Big Data
- 3. Quelles contre-mesures aujourd'hui ?

Le Big Data : vous pensez que c'est flou ? Vous avez raison !

Une multitude d'usages, des possibilités infinies... et beaucoup de buzzwords

Big Data: faire face au déluge de données

L'explosion des données n'est pas forcément un problème. Au contraire, le phénomène Big Data peut également être une opportunité.

SiliconANGLE » Super Bowl XLVIII Racks Up 'Super Big Data' Usage

Super Bowl XLVIII racks up 'Super Big Data' usage

MELLISA TOLENTINO | FEBRUARY 4TH

Big Data Paris 2014 : les groupes du Cac 40 comme grands témoins

« Big data », l'intelligence de l'énergie ?

si ceux de l'énergie, des transports, de la santé... touche. Défi pour les uns, solution miracle pour les autres. Défi pour ceux qui n'a jamais eu à se soucier de la maîtrise des raffinements du jargon anglo-américain.

Sanofi, GDF Suez... Lors du salon dédié au Big Data qui se tient les 1er et 2 avril au CNIT de Paris - La Défense. les plus grands groupes français de l'énergie.

BIG DATA MOBILE

Décollage du big data et des apps dans les compagnies aériennes

Les 3 V du Big Data : Volume, Vitesse et Variété

Le volume des données explose. Dans un rapport de 2013 consacré au Big Data, McKinsey prédisait une augmentation de 60 % de la marge d'exploitation des détaillants qui utiliseront le Big Data.

Facebook: 10 ans de big data !

Le dimanche 02 Mars 2014 à 12:10 par Jérôme G. | 2 commentaire(s)

Big Data et vie privée : 71% des Français inquiets quant à la collecte de leurs données

Par Marine Doffrennes
Publié le 25 février 2014

Oracle rachète le spécialiste du Big Data marketing BlueKai pour 400 millions de dollars (estimation)

Par Olivier Harmant | le 25 février 2014 | 0 Commentaire

Acquisition Actualité Big Data

HARVARD
MAGAZINE

NEWS RESEARCH STUDENTS ALUMNI ARTS SPORTS HARVARD

Plus > March-Avril 2014 Back Issues Harvard Finances Undergraduate

FEATURES

Why "Big Data" Is a Big Deal

Information science promises to change the world.

Le Big Data, c'est comme le sexe chez les ados

ACTUALITÉS

L'actu tech en Asie : la Chine, possible Eldorado du « big data »

Par Les Echos | 14/02 | 16:34 | mis à jour le 18/02 à 18:37

Healthcare Big Data in 3 Easy Steps

October 28, 2013 - by Shaun Whatley - in System

Big Data will change the way we practice, fund, and regulate medicine. Patients and clinicians need to know how this will impact them.



This post asks "What is Big Data?", "How Will It Help Healthcare?", and "What Should We

Le Big Data : vous pensez que c'est flou ?

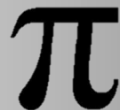
Vous avez raison !

Certains vont même jusqu'à parler d'un nouveau paradigme pour la science

Sciences
empiriques



Sciences
théoriques



Sciences
informatiques



?



Le Big Data : vous pensez que c'est flou ?
Vous avez raison !

**...Mais les métiers demandent
du Big Data !**

Le Big Data vous pensez que c'est flou ?

...Mais les métiers demandent du Big Data !

Parfois à tort...

- Tout ce qui est « volumineux » est Big Data
- Sirènes marketing et revival de la BI

« On parlait de Big Data car le stockage était le principal problème. Mais 6 ans plus tard, Hadoop et HBase permettent de préparer et de stocker de larges volumes de données, et le terme erroné Big Data persiste encore ».

« Aujourd'hui, les entreprises peuvent à peine parler de **medium data**, car ce qu'on appelle le Big Data aujourd'hui ce sont les méthodes analytiques, pas les données elles-mêmes ».

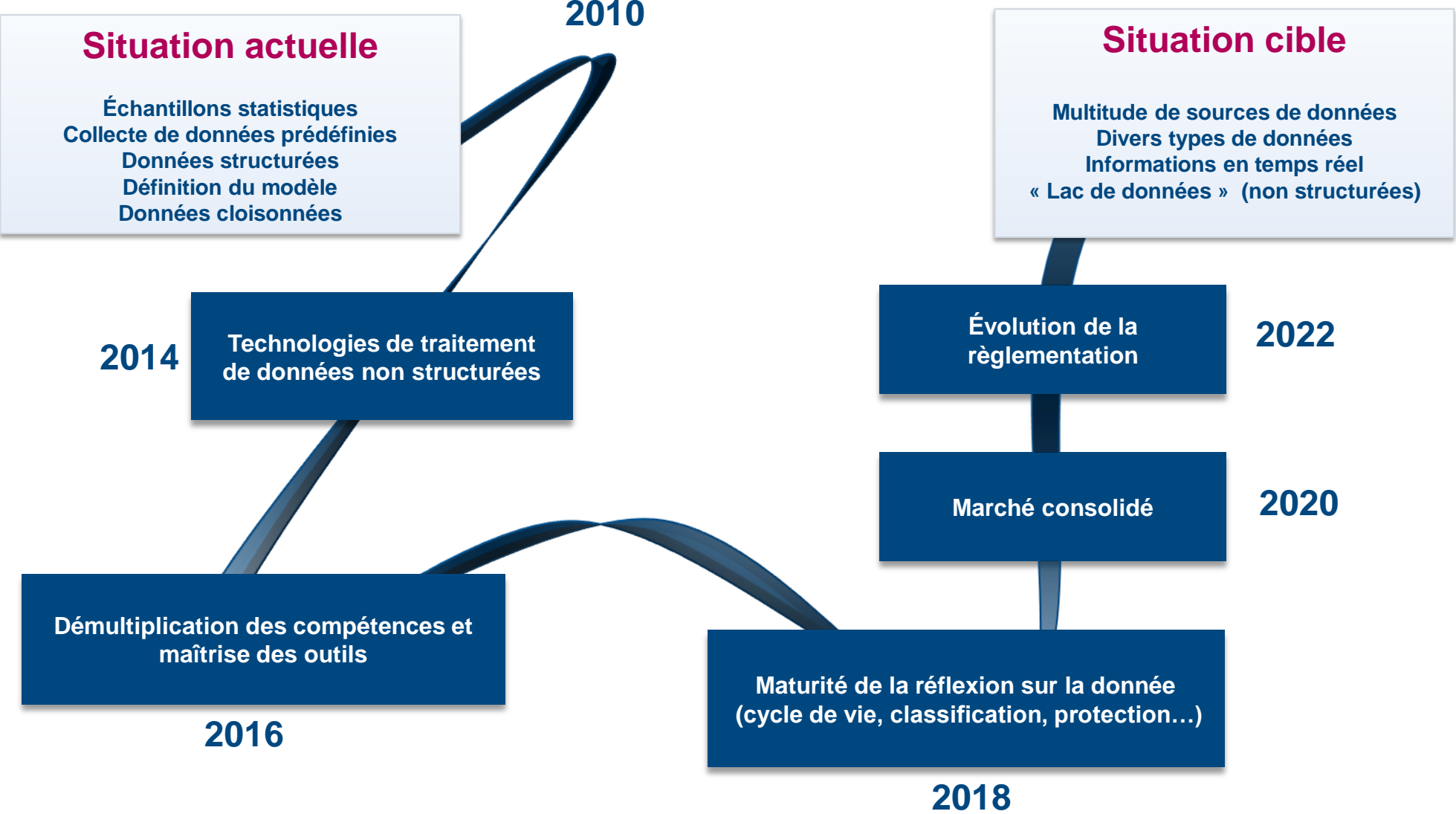
Harper Reed et Rayid Ghani,
(Directeurs IT de la campagne de Barack Obama en 2012)

... Mais aussi à raison :

- Augmentation indéniable des volumes et des types de données à traiter
- De réelles opportunités pour créer de la valeur

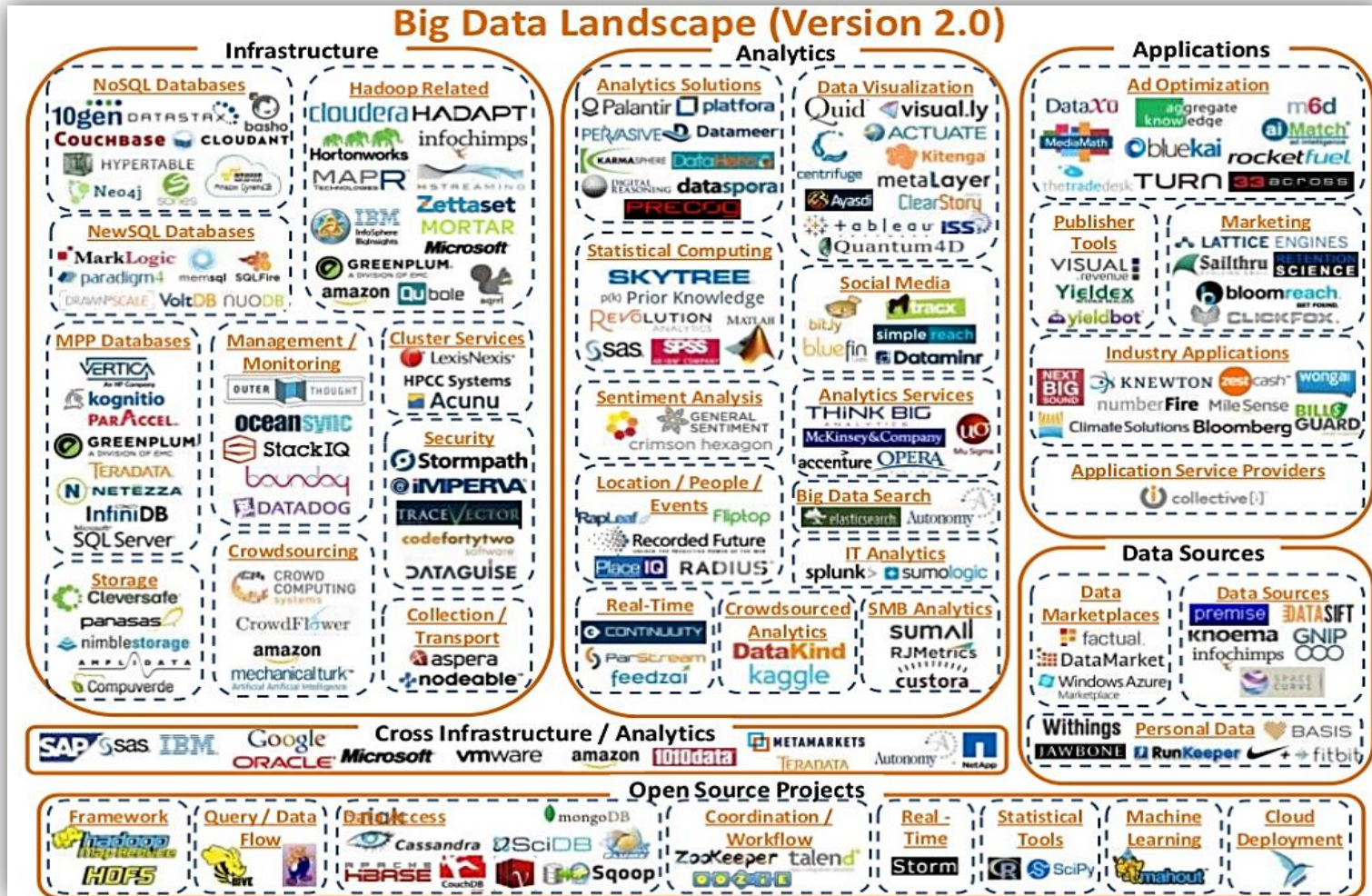


Le Big Data vous pensez que c'est flou ? Où en sommes nous ?

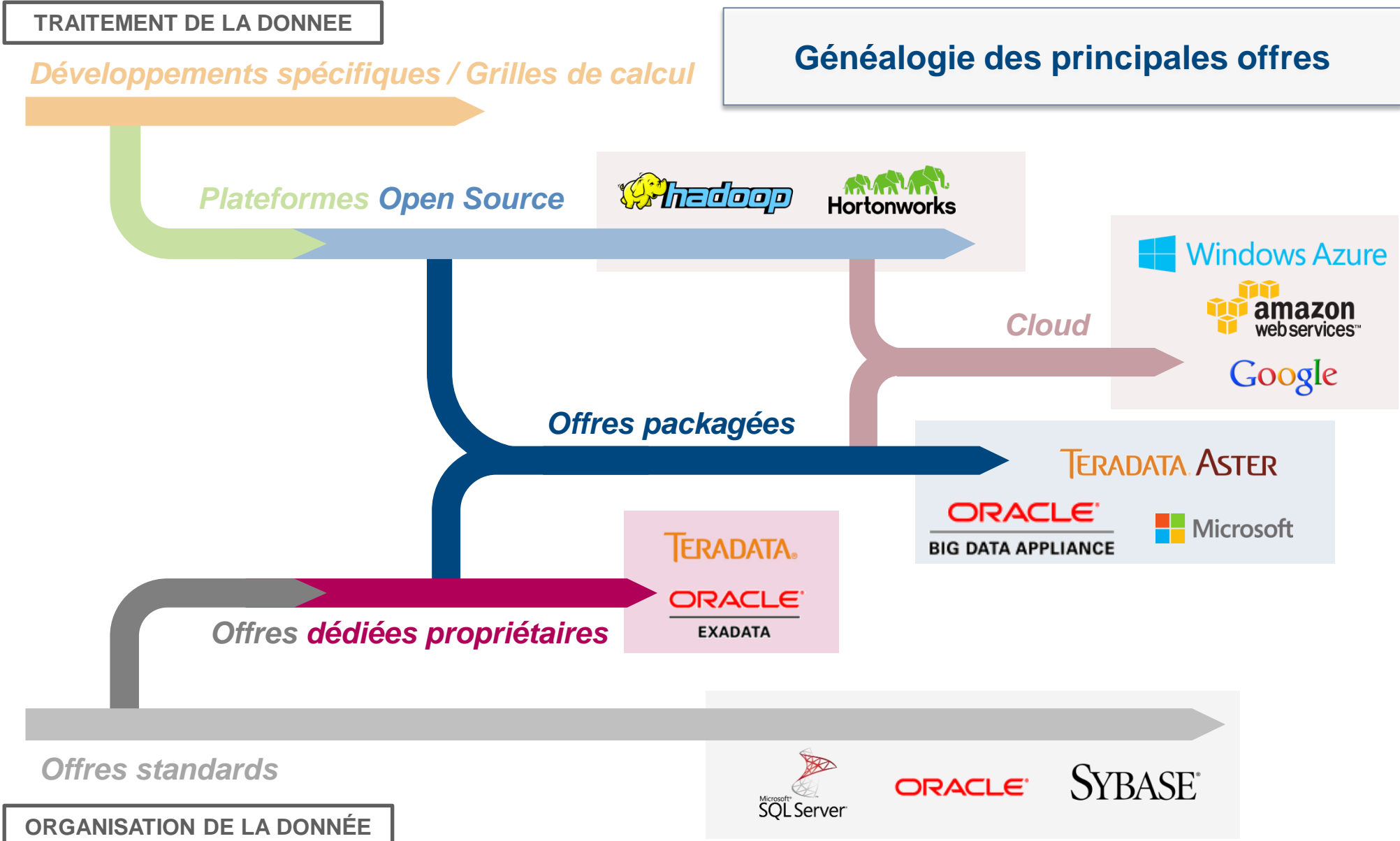


Le Big Data vous pensez que c'est flou ? Où en sommes nous ?

Une multitude de technologies (et plusieurs initiatives par semaine)



Le Big Data vous pensez que c'est flou ? Où en sommes nous ?



Le Big Data : vous pensez que c'est flou ?
Vous avez raison !

**Le Big Data qui marche c'est celui qui a
un apport métier !**

Le Big Data : vous pensez que c'est flou ?

Quelques exemples IRL

NSA



Volume

Variété

Vélocité

5 zettaoctets= 10^{12} Go
 =
 5.000.000.000.000.000.000.000 d'octets
 =
 5 000 milliards de milliards d'octets

Microsoft/INRIA



Volume

Variété

Vélocité

10⁶ images de cerveau
 10⁶ données génétiques
 x2000 types différents

 Pixels 3D

 106 variables

Crédit Mutuel Arkea



Volume

Variété

Vélocité

Ratios de solvabilité des offres de prêts
 48 heures → 15 minutes

 Lutte anti blanchiment
 6 mois d'historique → 20 secondes

1. Le Big Data vous pensez que c'est flou?
- ▶ **2. Les grands risques du Big Data**
3. Quelles contre-mesures aujourd'hui ?

Les grands risques du Big Data

Les risques principaux sont liés à la donnée...

Sources de données

Multiplication et appartenance des sources de données



Régulation de la donnée

Données à caractère personnel mais pas uniquement



Accès et cycle de vie de la donnée

Droit d'accès et cycle de vie des données



Les grands risques du Big Data

... et se positionnent sur les étapes du cycle de vie

Sources de données

Régulation de la donnée

Accès et cycle de vie de la donnée

1 Inclusion de données de mauvaise qualité ou malicieuses
(I)

2 Acquisition / Inclusion de données interdites au regard de la législation / des contrats
(CR)

3 Traitements non conformes (en particulier désanonymisation par inférence)
(CR)

4 Accès illégitime aux données et traitements
(CR)

5 Difficulté pour identifier / modifier des données stockées
(CR)

6 Publication illégitime / malicieuse de données
(CR)

Technologiques

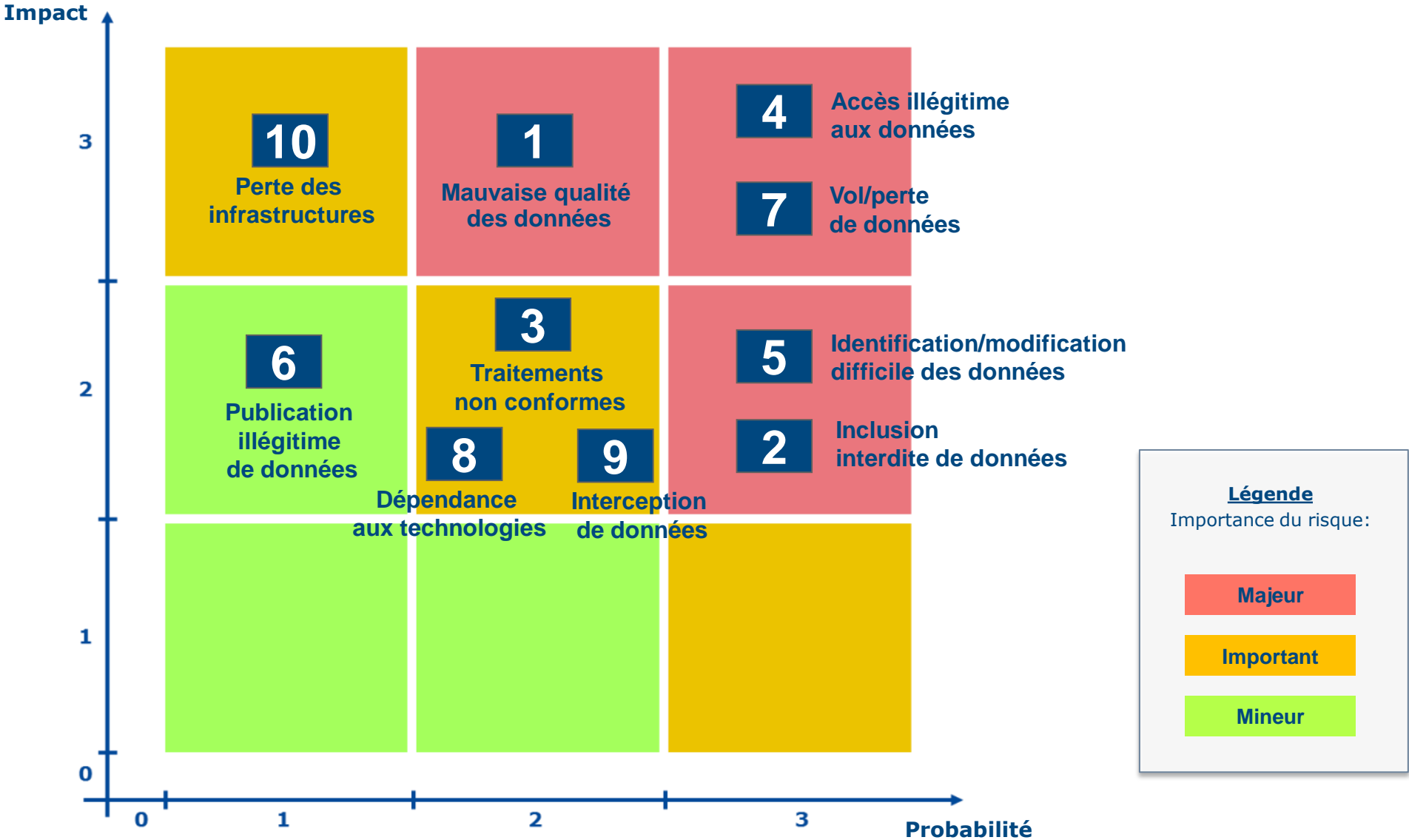
7 Vol/perte de données à cause d'une mauvaise maîtrise des nouvelles technologies
(CIDR)

8 Dépendance à des protocoles, des applications ou des technologies jeunes et mouvantes
(D)

9 Interception de données (CI)
10 Perte des infrastructures ou bases de données (D)

Les grands risques du Big Data

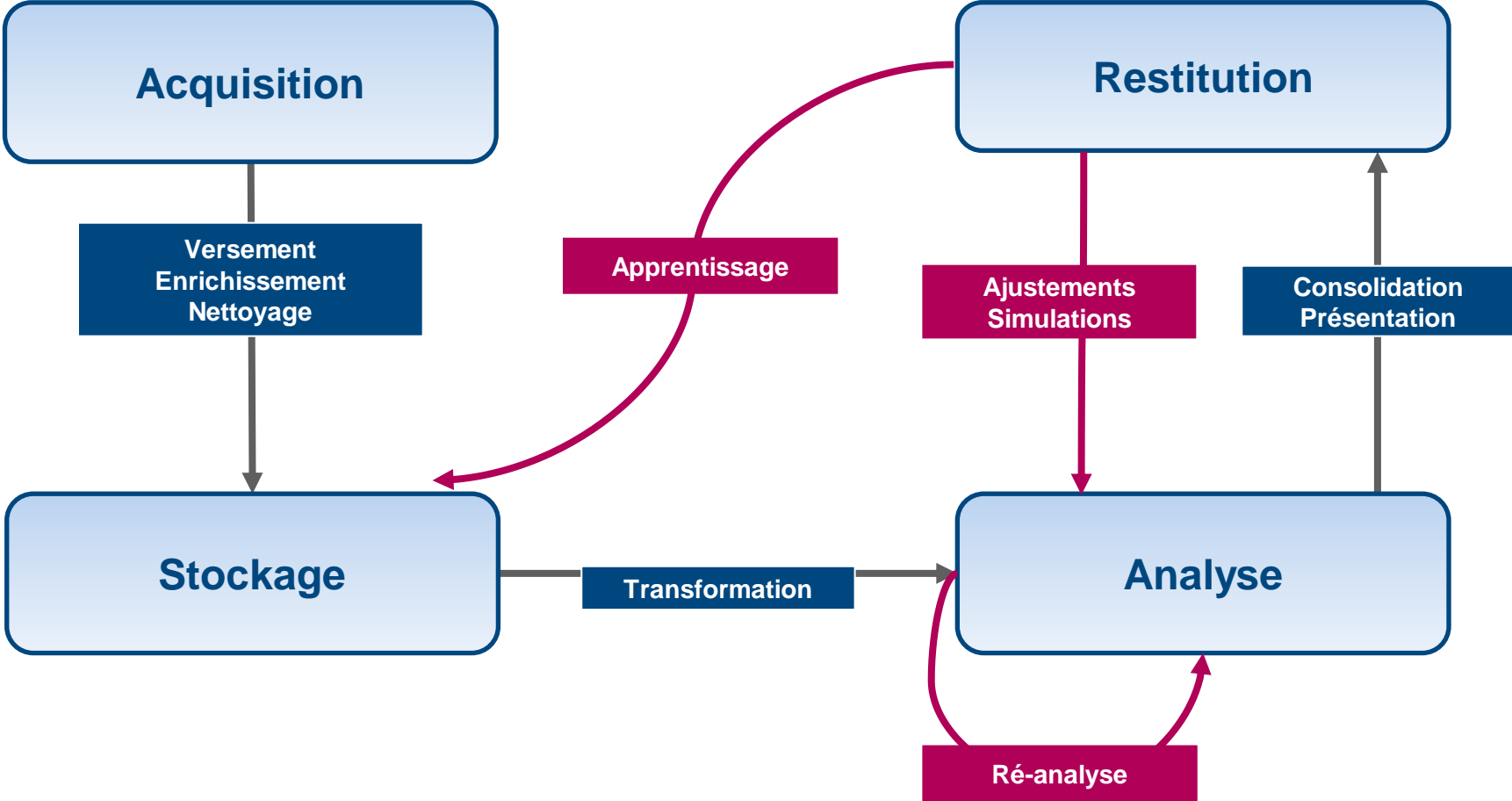
Tout dépend du contexte, mais des tendances se dessinent



1. Le Big Data vous pensez que c'est flou?
2. Les grands risques du Big Data
- ▶ 3. Quelles contre-mesures aujourd'hui ?

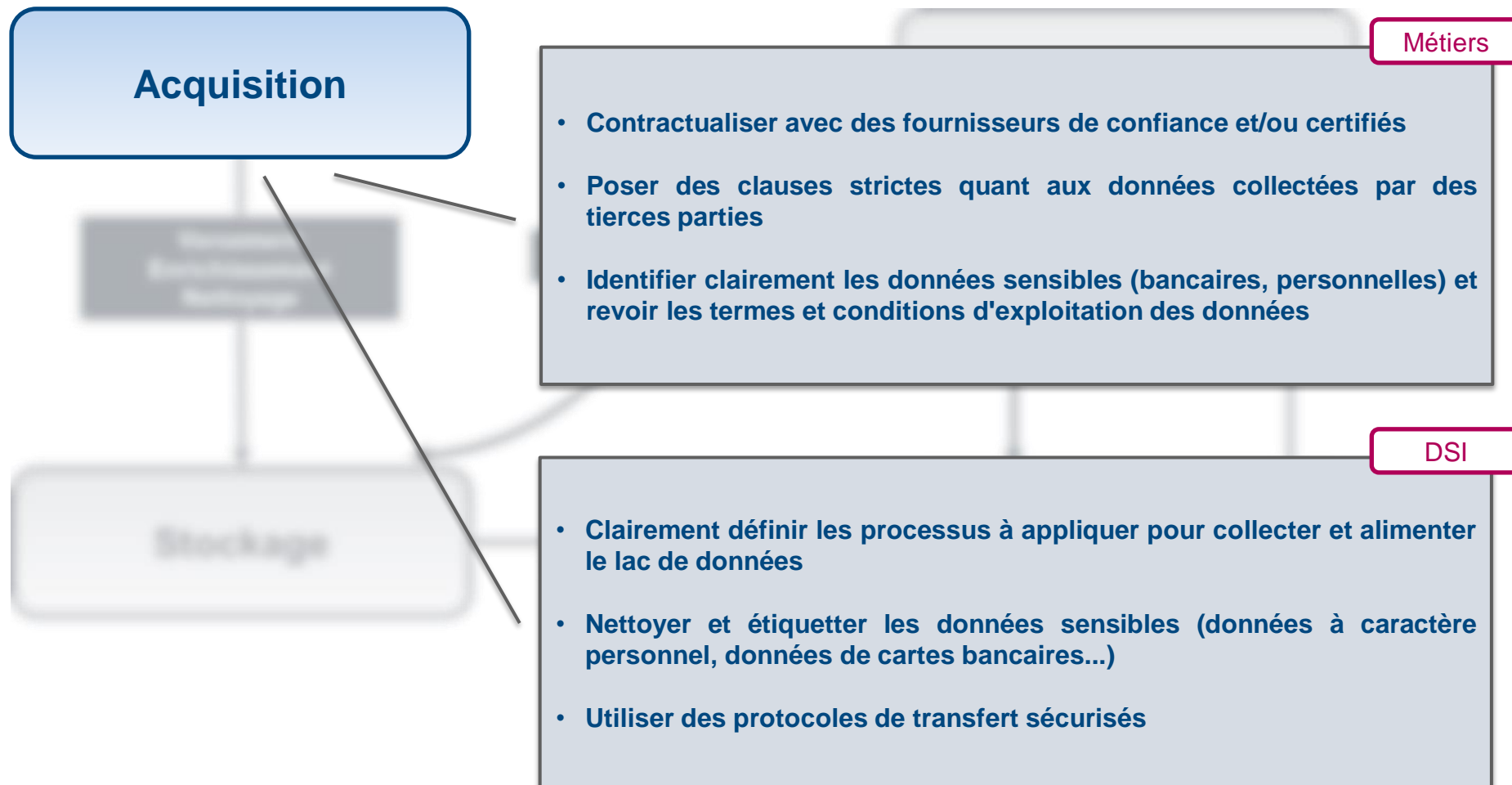
Quelles contre-mesures aujourd'hui ?

Des contre-mesures à déployer tout au long du cycle de vie



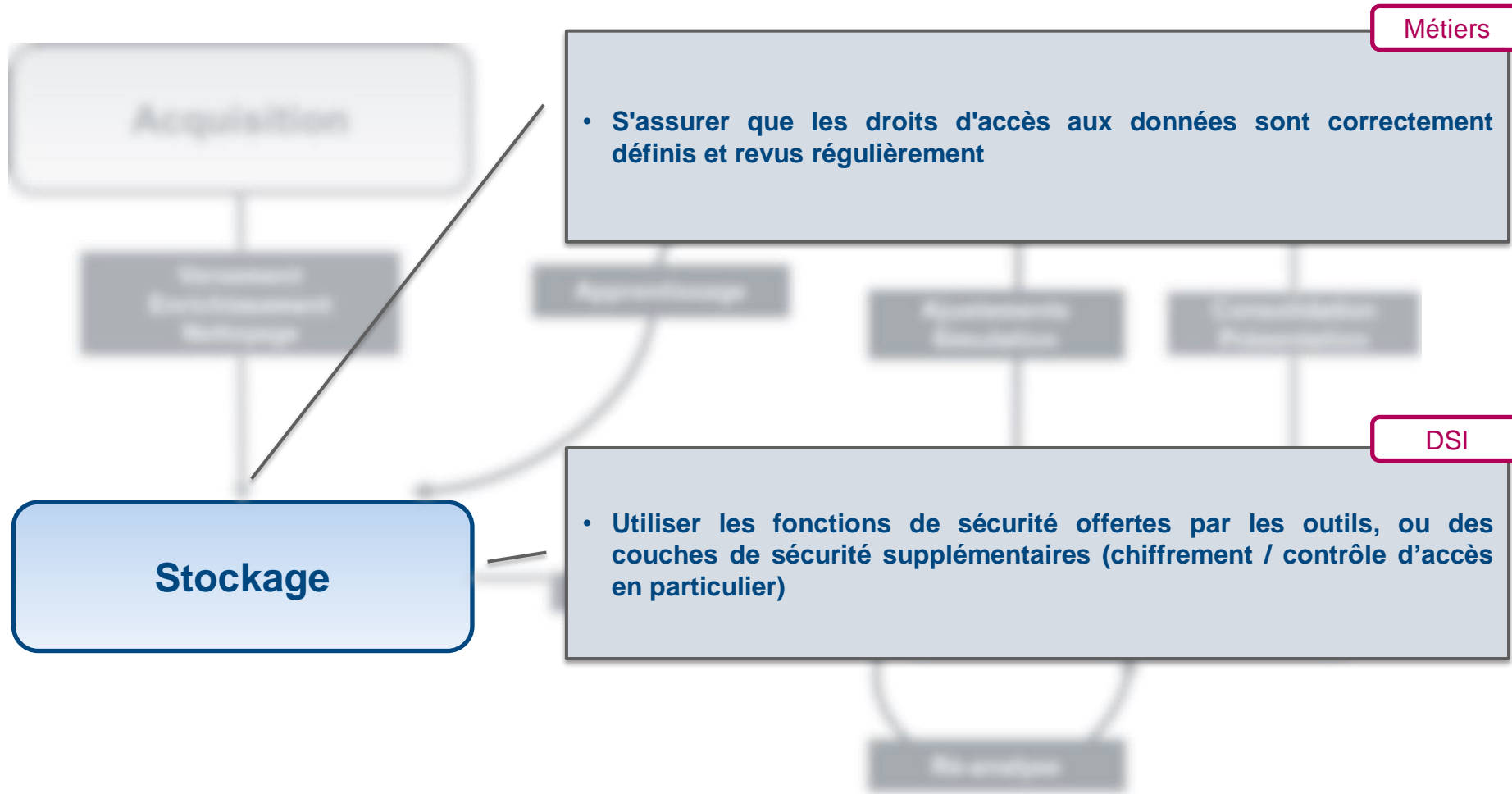
Quelles contre-mesures aujourd'hui ?

Des contre-mesures à déployer tout au long du cycle de vie



Quelles contre-mesures aujourd'hui ?

Des contre-mesures à déployer tout au long du cycle de vie



Quelles contre-mesures aujourd'hui ?

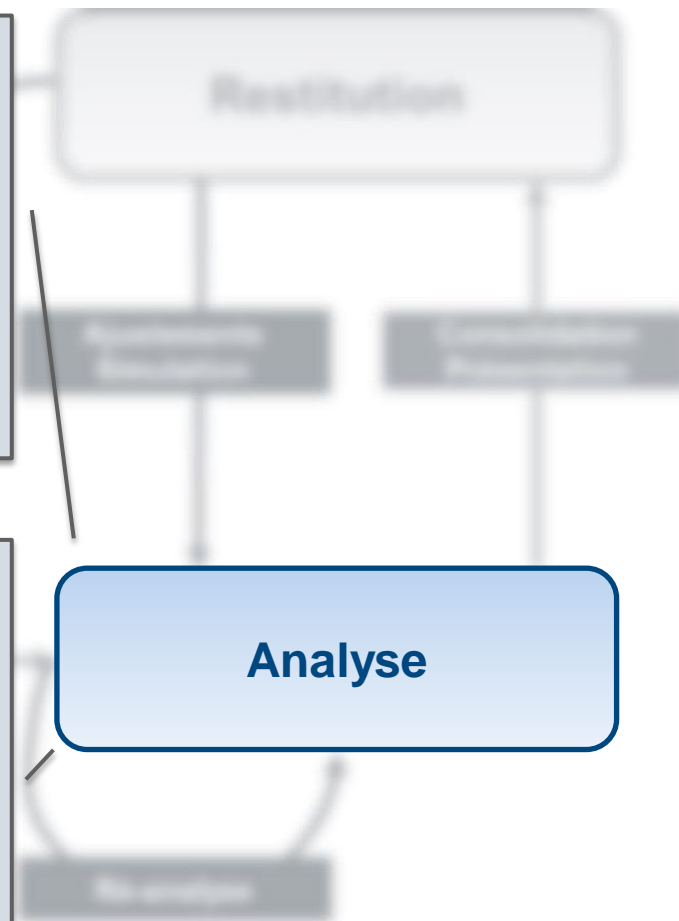
Des contre-mesures à déployer tout au long du cycle de vie

Métiers

- Définir clairement les traitements appliqués et les buts de l'analyse
- Déclarer les traitements de données aux autorités/personnes en charge de la protection des données
- Définir les modèles et réaliser les analyses de données avec des Data Scientists (spécialistes de l'analyse statistique)

DSI

- Anonymiser/masquer les données sensibles issues des traitements
- Mettre en place des mesures techniques pour clairement délimiter le domaine de l'analyse
- Mettre en place des contrôles d'accès aux données et aux traitements



Quelles contre-mesures aujourd'hui ?

Des contre-mesures à déployer tout au long du cycle de vie

Métiers

- Former régulièrement les équipes à l'utilisation des nouveaux outils
- Établir des processus de contrôle des analyses Big Data (limites, portées...)

Restitution

DSI

- Réaliser des audits réguliers et insérer le Big Data dans les processus sécurité existants (patch management, durcissement, gestion de l'administration, DRP...)
- En capitalisant sur les approches Cloud si c'est l'orientation choisie !

Quelles contre-mesures aujourd'hui ?

Quelle approche pour la sécurité des systèmes Big Data ?

- Un **besoin criant** de fonctions de sécurité
 - Contrôle d'accès, chiffrement, anonymisation/tokenization, traçabilité, transfert de données...
- Une **maturité très variable des technologies Big Data**
 - Un sujet (souvent) oublié à la conception des solutions...
- Les **approches existantes atteignent leur limite**
 - D'une approche centralisée... à une décentralisation
 - De silos de données maîtrisées... à des lacs de données variées

Aujourd'hui un marché en pleine ébullition !

Quelles contre-mesures aujourd'hui ?

Quelle approche pour la sécurité des systèmes Big Data ?

Vers la mise en place de mesures complémentaires

Quelles solutions ?

- Systèmes centralisés
- IAM "classique"
- Chiffrement intégral



- Systèmes décentralisés (et parfois dans le Cloud)
- Une approche contrôle d'accès par "attribut" lié à chaque cellule (ABAC)
- Chiffrement par attribut / homomorphe

Quelle cible ?



Infrastructure



Base



Donnée

Quels acteurs ?



Zettaset

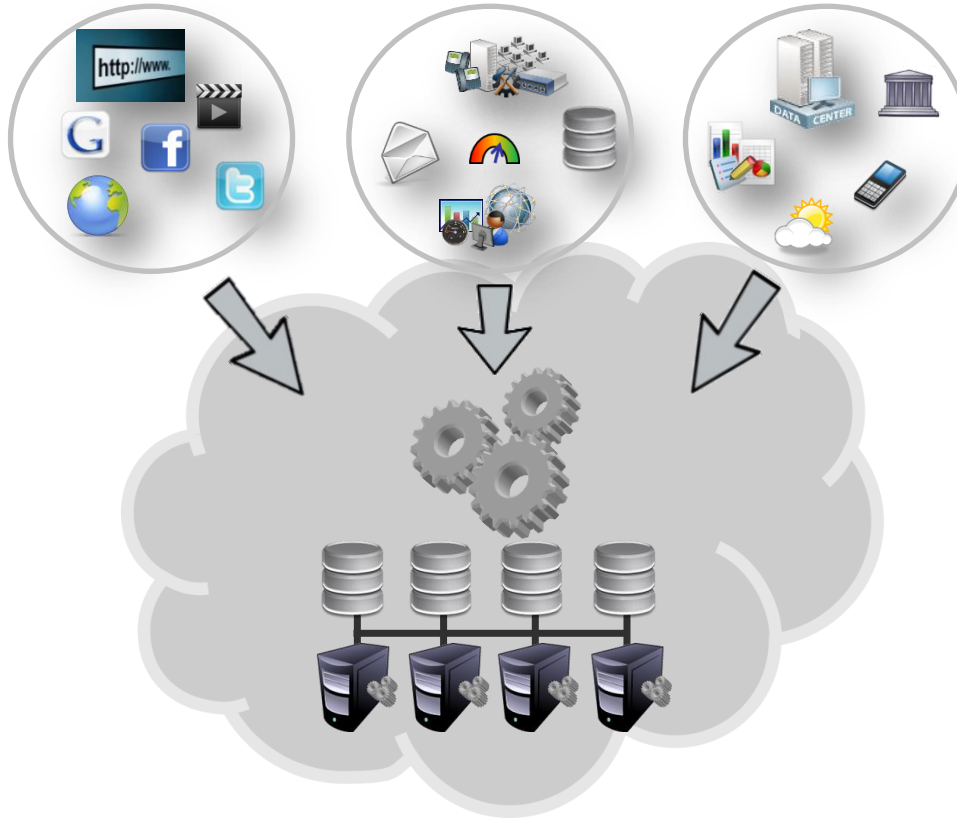
cloudera



Mais en restant attentif à l'augmentation du niveau de complexité...

Les questions à se poser pour démarrer !

- A qui **appartiennent les données** que je collecte ?
- Quelle est **leur nature** ? (données spécifiques, données personnelles...)



- Qui sont mes **fournisseurs externes** et quelle est la structure de mon **contrat** avec eux ?

- Quel est mon **objectif** vis-à-vis de ces données... et est-il **déclaré** ?
- Existe-t-il des **contraintes réglementaires** particulières ?

- Quels **types de traitement** vont être réalisés et par qui ?
- Les résultats seront-ils **utilisés légitimement** ?

- Sais-je **où les données sont stockées** et suis-je capable de les **modifier** ?
- **Combien de temps** seront-elles stockées ?

- Les technologies sont-elles **maîtrisées** ?
- Sont-elles utilisées dans des versions **à jour** ?
 - Sont-elles entièrement **maintenues** ?

- Chaque type de donnée est-il **protégé de manière adéquate** contre les différentes menaces ? (modification, accès illégitimes, vol...)



www.solucom.fr

Contacts

Gérôme BILLOIS

Senior Manager

gerome.billois@solucom.fr

Twitter : @gbillois

Chadi HANTOUCHE

Manager

chadi.hantouche@solucom.fr

Twitter : @chadihantouche