

# Les particularités des projets liés au « Big Data »

Emmanuel Faug

Directeur Services-conseils en intelligence d'affaires chez Momentum Technologies

Le 8 décembre 2015

# Une définition simple du « Big Data »

- L'expression « **Big Data** » (traduite en français par « mégadonnées » ou « données massives ») désigne la **masse hétérogène** des données numériques produites par les entreprises et les particuliers dont les caractéristiques (très grand volume, diversité de forme, vitesse de traitement) requièrent des **outils d'analyse informatiques spécifiques**.

*Source - Définition « Le Crédit Lyonnais » (LCL)*

- Le cerveau humain ne peut enregistrer qu'une quantité limitée d'information et ne peut donc appréhender certaines analyses trop complexes

*Source - consoGlobe*

# Une définition simple du « Big Data »

- Les 3 « V »... et même les 6 « V »...



# Peu d'initiatives... et beaucoup d'échecs...

1 « Près des **deux tiers** des projets Big Data ne franchiront pas le stade du pilote et de l'expérimentation au cours des deux prochaines années, et finiront par être abandonnés. »

*Source - Vos projets Big Data échoueront probablement. Voici pourquoi  
Par Steve Ranger, ZDNet.com  
Jeudi 17 Septembre 2015 (Gardner)*

2



*Source - Big Data : peu de projets, le plus souvent des échecs  
Par Christophe Auffray, ZDNet.com  
Mardi 20 Janvier 2015 (Capgemini)*

# Raisons multiples de l'échec

- Manque de vision
  - cas d'usage mal voire non identifié
- Perte du focus sur le cas d'usage initial : éparpillement
- Conduite du changement négligée voire occultée
- Contraintes organisationnelles trop fortes
- Approche sous l'angle technologique (infrastructures et logiciels)
  - « bruit » marketing des fournisseurs de solutions « Big Data »
- Manque d'expertise
- Manque de motivation



# Exemples de projets liés au « Big Data »

- Quelques exemples significatifs dans des domaines variés
  - Vente en ligne, banque et assurance
    - **Connaissance et expérience client** : analyse du comportement du client sur le Web (site de la banque, réseaux sociaux, recherches réalisées sur le Web), des données internes d'historique du client pour adapter l'offre personnalisée au client
    - **Lutte contre la fraude bancaire** : croisement des informations reliées à une transaction de paiement avec l'historique des transactions du client, son profil d'acheteur, ses activités sur les réseaux sociaux (pour le géolocaliser par exemple)
  - Transport
    - **Optimisation des infrastructures et les services en ville** : modéliser les déplacements des gens afin d'adapter les infrastructures et les services (données provenant des cartes de transport en commun, géolocalisation des personnes et des automobiles)
    - **Optimisation des vols aériens** : analyser les données en vol pour optimiser les couloirs aériens et réaliser des économies de carburant (données provenant des capteurs sur les avions et de la météo)

# Exemples de projets liés au « Big Data »

- Quelques exemples significatifs dans des domaines variés
  - Industrie manufacturière : chaîne de production
    - **Maintenance préventive pour anticiper les pannes** : analyse des données provenant des capteurs sur les machines
  - Gouvernement : déclaration de revenus
    - **Lutte contre l'évasion fiscale** : croisement de données provenant d'une multitude de sources d'informations sur le contribuable (particulier ou entreprise) afin de détecter des comportements frauduleux vis-à-vis des déclaration de revenus
  - Santé
    - **Médecine préventive et personnalisée** : analyse des données des patients provenant de leurs dossiers médicaux et d'autres sources d'information (objets connectés, recherches effectuées sur le Web, réseaux sociaux) pour faire de la prévention, réduire les coûts de santé et ouvrir la voix à la médecine personnalisée

# Exemples de projets liés au « Big Data »

- Quelques exemples significatifs dans des domaines variés
  - Environnement
    - Identification du lieu idéal pour placer une éolienne
  - Télécommunications
    - Prévoir et prédire le taux de désabonnement, promouvoir la fidélisation
  - Politique
    - Optimisation d'une campagne électorale : analyser des données du Web pour cibler les indécis et les thèmes sur lesquels il était possible d'influencer leur vote afin de leur adresser les bons messages électoraux (exemple : cas de l'élection de Barack Obama aux États-Unis en 2012 et des « swing states »)



# Caractéristiques générales d'un tel projet

- **Fort degré d'incertitude et dimension exploratoire !!!**



- **Transversalité et innovation**

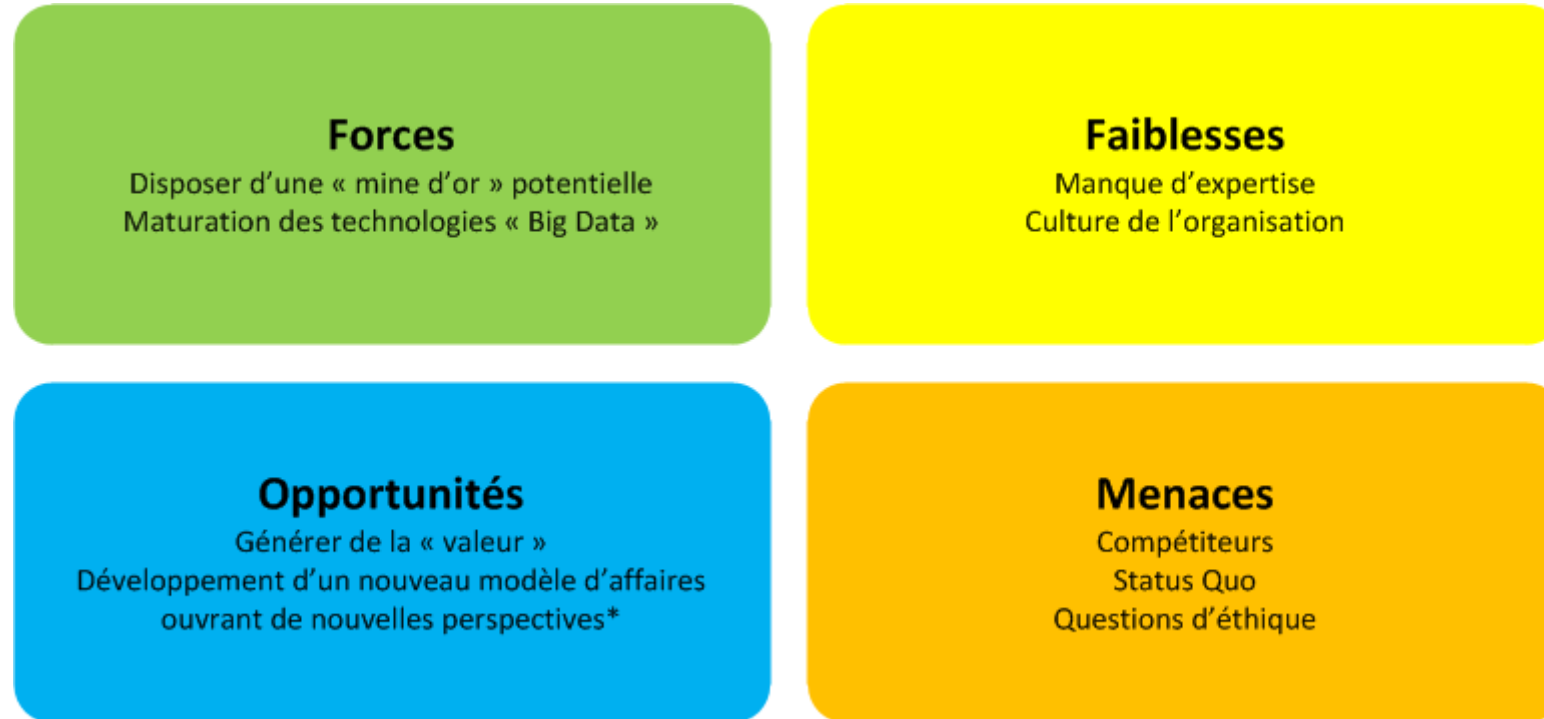
# Caractéristiques générales d'un tel projet

- Paradoxe

- Une voix nous dit « Il faut y aller ! »
  - *On est tellement convaincu qu'il existe de la valeur dans ces données massives*
- Une autre nous dit « Il ne faut pas y aller ! »
  - *On n'est tellement pas certain du résultat : distinguer le « signal » du « bruit » !*



# Enjeux (matrice de SWOT)



\* Perspectives non seulement financières mais aussi sociétales, sociales, environnementales, etc.

# Défi scientifique

- Compétence clé (et plutôt rare) en science de la donnée
- Scientifique de la donnée
  - Expertise en modélisation et en mathématiques, en construction d'algorithmes « déductifs » mais aussi et surtout « inductifs » (découverte et mise en évidence de corrélations)
  - Expertise en visualisation de la donnée pour découvrir des patrons (« patterns ») dans les données permettant de dégager de la valeur pour l'organisation

# Défi technologique



# Défi technologique

- Compétences TI spécifiques
  - Spécialistes en infrastructures et services de collecte, d'intégration et d'exploitation de données massives (pouvant adresser tous les « V »)
- Technologie « Big Data » = technologie disruptive
  - Stockage, traitement et exploitation de l'information :
    - Des modèles de fichiers plats... aux modèles de données relationnels (SGBDR et SQL)... puis aux modèles de données NoSQL (clé-valeur, colonnes, documents, graphes)
    - Visualisation avancée des données
  - Notion d'écosystème « ouvert » en évolution « forte » et « rapide » par rapport à des systèmes « propriétaires » plutôt monolithiques

# Défi organisationnel

- Projet « transverse »
- Organisation « fluide »
  - « Affaires » et « TI » : faire émerger un nouveau modèle de collaboration renforçant la synergie entre les deux lignes
  - **Entreprise ouverte et organisée en réseau** plutôt que fermée et organisée en silos
- Culture et gouvernance de la donnée, transformation digitale
  - Développer la **culture de la donnée** au sein de l'entreprise, comprendre le potentiel de la donnée
  - **Faciliter l'accès à la donnée** tout en gérant les aspects de sécurité, de confidentialité et d'éthique engendrés par l'exploitation des données massives (résoudre les dilemmes, trouver le bon compromis)

# Défi humain

- Éléments critiques à prendre en considération
  - **Motivation**, **Imagination** et **Esprit d'innovation** vs **Expertise**
  - L'**Expertise** qui est indispensable au sein de l'organisation se doit d'être très évolutive : elle peut s'acquérir par la collaboration avec des partenaires extérieurs à l'organisation, par les activités de veille et de formation à l'interne...
  - ... tandis que la **Motivation**, l'**Imagination** et l'**Esprit d'innovation** sont des éléments tout aussi indispensables mais bien plus délicats à faire évoluer



# Défi culturel

- Transformation culturelle
  - Faire rentrer l'esprit d'une startup dans l'organisation

## INTRAPRENEURIAT

<https://fr.wikipedia.org/wiki/Intrapreneuriat>

### Extrait de l'article Wikipedia

« Le concept a été défini en 1976 par l'américain [Ginford Pinchot](#). L'intrapreneuriat est censé permettre à la grande entreprise de mieux saisir les opportunités que son inertie naturelle laisserait passer sinon. »

# Approche de mise en œuvre

- 3 champs d'expertise nécessaires et suffisants



# Approche de mise en œuvre

- Monter une « petite » cellule d'innovation (**INTRAPRENEURIAT**)
  - Mêlant les lignes « Affaires » et « TI »
  - Fonctionnant en mode « agile »
- Composition : un triptyque « Sciences - Affaires - TI »



# Approche de mise en œuvre

- Rôle de chacun dans l'équation
  - « Affaires » : établir la vision et les objectifs, les suivre
    - Identifier la « valeur potentielle » de l'information : le **cas d'usage**
  - « Sciences de la donnée » : tenir compte des objectifs d'affaires
    - Traiter les enjeux liés à l'analyse des données et les techniques et outils les mieux appropriés
    - Traiter la véracité, la variabilité et aller chercher la valeur : **élaboration des modèles mathématiques d'analyse et conception des algorithmes (analytique avancée)**
  - « TI » : tenir compte des besoins des scientifiques de la donnée
    - Traiter les enjeux liés à la technologie « Big Data » et les infrastructures, les techniques et outils les mieux appropriés pour pouvoir gérer le volume de données, la variété des sources et la vélocité de traitement

# Approche de mise en œuvre

« FAIL FAST, LEARN FAST ! »

1. Phase d'exploration initiale avec la cellule d'innovation
2. Phase de généralisation à l'échelle de l'organisation  
(industrialisation de la démarche « Big Data » au sein de l'organisation)

# Approche : phase initiale

- Voir grand
  - Vision
  - Innovation
  - Prise de risque
- Commencer petit
  - Identifier un **cas d'usage** « pilote » en s'appuyant par exemple sur la matrice de SWOT et monter la cellule d'innovation
  - Réaliser une **preuve de concept**
    - Être agile : mitiger le risque en validant au fur et à mesure les résultats obtenus vis-à-vis du cas d'usage identifié au départ
    - Se réajuster à chaque étape (mode « essai-erreur »)
    - Privilégier des cycles très courts à chaque itération

# Approche : phase de généralisation

*Source - Comment faire du Big Data une réalité, par Laurence Chrétien, Harvard Business Review, juillet 2015*

- Extraits résumant l'article
  - *Freins à la mise en œuvre d'un projet « Big Data » au sein d'une organisation*
    - *Freins majeurs : la dispersion des données et le manque de coordination*
    - *Autres obstacles : une gouvernance immature de la donnée et la dépendance aux systèmes hérités*
  - *Comment se distinguent les organisations qui y parviennent ?*
    - *Elles se dotent d'une véritable organisation « Big Data »*
    - *Elles nomment un Chief Data Officer (CDO)*
    - *Elles activent des leviers multiples pour monter en compétence*
  - *4 pièces essentielles dans le casse-tête*

Qualité du modèle  
opérationnel

Expérimentation (PoC)  
« fail fast, learn fast »

Alignement des « affaires »  
et des TI sur le potentiel du  
« Big Data »

Politique de sécurité des  
données draconienne

Merci !



[emmanuel.faug@momentum-tech.ca](mailto:emmanuel.faug@momentum-tech.ca)



# Références

- **LCL**
  - <https://www.lcl.com/guides-pratiques/zooms-economiques/big-data-banque.jsp>
- **consoGlob**
  - <http://www.consoglobe.com/big-data-et-environnement-cg>
- **ZDNET**
  - <http://www.zdnet.fr/actualites/vos-projets-big-data-echoueront-probablement-voici-pourquoi-39825066.htm>
  - <http://www.zdnet.fr/actualites/big-data-peu-de-projets-le-plus-souvent-des-echecs-39813301.htm>
- **Les Echos**
  - <http://www.lesechos.fr/idees-debats/cercle/cercle-86411-big-data-la-realite-des-projets-au-dela-du-buzz-1001055.php>
- **Harvard Business Review**
  - <http://www.hbrfrance.fr/chroniques-experts/2015/07/7818-comment-faire-du-big-data-une-realite/>