



neXialog  
CONSULTING

THINK SMART X ACT DIFFERENT

# Big Data, ce n'est pas un projet, c'est toute une démarche...

## Introduction

---

Depuis la révolution numérique, la masse de données produites chaque jour a pris de telles proportions qu'il est difficile de continuer à utiliser les outils traditionnels pour les manipuler de façon performante.

La plupart des entreprises se contentent d'exploiter une part réduite des informations générées et récoltées dans leur SI pour produire toujours les mêmes rapports. Pourtant, les possibilités ouvertes par le mouvement Big Data sont alléchantes.

La démarche Big Data est généralement initiée lorsque le management se rend compte que l'entreprise est en train de perdre des opportunités en négligeant les données à sa disposition. Elle semble primordiale si au moins une de ces affirmations est vraie :

- Son Système d'Information est complexe, étendu, composé de plusieurs briques logicielles.
- Son secteur d'activité est caractérisé par une production d'information en temps réel par une multitude de capteurs, d'utilisateurs ou de clients.
- Ses bases de données ne suffisent plus pour obtenir des résultats dans des délais raisonnables malgré les efforts d'optimisations menés à ce propos.
- Ses besoins de puissance de calcul peuvent varier de façon critique.
- Ses données ne sont pas forcément structurées et sont volumineuses.
- Ses données sont "qualifiées" ou méta-taguées de façon disparate.

Mais comme le Big Data, par définition, nous soumet sans cesse de nouvelles variétés de données, toujours plus vite et avec des volumes croissants, les entreprises peuvent être confrontées en permanence à des inconnues dans leur équation et avoir sans cesse à résoudre des problèmes nouveaux.

### **Mots clés :**

Big Data, Volume, Variété, Vitesse, Vérité, ETL, Hadoop, Olap, Data scientists.

## Contenu

---

Contenu.....	2
Définition .....	4
Big Data, des cas d'usages bien multiples.....	5
Cinq actions pour mettre en place une démarche Big Data.....	8
Faire l'inventaire de ses données .....	8
Donner la priorité aux métiers .....	8
Réévaluer l'infrastructure .....	9
Etudier les technologies associées.....	9
Préparer ses équipes .....	9
Les compétences nécessaires pour une démarche Big Data .....	10
Data scientists.....	10
Spécialistes des outils d'ETL.....	10
Développeurs Hadoop.....	10
Spécialistes des outils de visualisation .....	10
Développeurs OLAP .....	10
Pourquoi le big data ne tient pas encore ses promesses.....	11
Les compétences nécessaires sont encore rares.....	11
Les responsables opérationnels restent difficiles à convaincre .....	11
Toutes les données ne sont pas exploitables.....	12
La rentabilité des projets n'est pas toujours au rendez-vous.....	12
L'offre technique se révèle pléthorique et complexe.....	12
Focus, Big Data et le secteur bancaire.....	13
Introduction .....	13
Big Data, une fonction Risques non isolée .....	13
Big Data, une fonction Risques organisée et intelligente.....	15
Big Data, des process de production simples .....	16
Anticipation sur l'évolution d'une fonction Risques (en environnement Big Data).....	17
Focus, Big Data et le secteur assurantiel.....	<b>20</b>
Principales opportunités pour l'assurance : .....	20
Zoom, Big Data dans la tarification en assurance : .....	20
Big Data et cadre réglementaire.....	22
Conclusion.....	23

## Définition

---

Le Big Data est dessiné pour offrir un accès et une exploitation en temps réel à des bases de données (de différents types) géantes. Il vise à proposer une alternative aux solutions traditionnelles de bases de données et d'analyse (serveur SQL, plate-forme de Business Intelligence...). Confronté très tôt à des problématiques de très gros volumes, les géants du Web, au premier rang desquels Yahoo (mais aussi Google et Facebook), ont été les tous premiers à déployer ce type de technologies. Selon le Gartner, le Big Data (en français "Grandes données") regroupe une famille d'outils qui répondent à une triple problématique (la règle dite des 3V):

- un **Volume** de données important à traiter; c'est la capacité technologique à effectivement récupérer, centraliser, stocker et requêter ces données dans des environnements sécurisés et à les mettre à disposition de façon utilisable
- une grande **Variété** d'informations (en provenance de plusieurs sources, non-structurées, structurées, Open...); puisque les données sont différentes, plus nombreuses et d'origines plus diverses, elles doivent permettre d'obtenir des informations inédites.
- un certain niveau de **Vélocité** à atteindre; c'est-à-dire l'exploitation de ces données requiert une grande Vélocité, par un lien dynamique en temps réel entre l'information produite et l'action opérationnelle qui en découle.

Bien que ces trois dimensions couvrent les attributs clés du Big Data, les organisations devraient prendre en compte un quatrième élément : la **Véracité**. La prise en compte de la véracité comme quatrième attribut des Big Data souligne l'importance de traiter et de gérer l'incertitude<sup>1</sup> inhérente à certains types de données (*les sentiments et la sincérité pour les humains, les conditions météorologiques, les facteurs économiques, ou encore l'avenir. comme exemple*).

La véracité fait référence au niveau de fiabilité associé à certains types de données. Chercher un niveau de qualité élevé est une exigence clé et un défi majeur des Big Data, mais même la meilleure méthode de nettoyage des données ne peut pas supprimer le caractère aléatoire inhérent à certaines données comme le temps, l'économie ou les décisions d'achat futures d'un client. Le besoin de reconnaître et de planifier cette incertitude est une dimension des Big Data qui est apparue lorsque les dirigeants ont cherché à mieux comprendre la nature incertaine de ces données.

---

<sup>1</sup> *Un exemple de cette incertitude concerne la production d'énergie : Le temps est incertain mais une compagnie d'électricité doit être capable de prévoir la production. Dans de nombreux pays, les régulateurs exigent qu'un pourcentage de la production provienne de sources renouvelables, mais ni le vent ni les nuages ne peuvent se prévoir avec précision. Alors comment pouvez-vous planifier ?*

*Pour gérer cette incertitude, les analystes doivent créer un contexte autour des données. Une façon de créer ce contexte consiste à fusionner les données afin que la combinaison de sources multiples peu fiables produise un point de données plus précis et plus utile, comme les commentaires sociaux ajoutés aux informations de localisation géo spatiale. Une autre manière de gérer ce caractère aléatoire peut être l'utilisation de mathématiques avancées pour couvrir cette incertitude, comme les techniques d'optimisation.*

## Big Data, des cas d'usages bien multiples

Le Big Data est un phénomène émergent. Pourtant, des cas d'usage courants sont déjà connus et apportent dès à présent une valeur significative.

### **Services Financiers (Banque, Assurance, Gestion de l'investissement)**

- *Conformité réglementaire,*
- *Gestion des risques,*
- *Prévention et détection des fraudes,*
- *Gestion des réclamations*
- *Optimisation de portefeuilles produits et clients*

### **Industrie Pharmaceutique**

- *Informations patients, télésurveillance*
- *Modèles prédictifs de santé, données embarquées*
- *Optimisation vente & distribution des médicaments*
- *Suivi des problèmes*
- *Gestion réglementaire*

### **Loisirs, Media & Communication - Retail Technology - Energie & Utilities**

- *E-marketing pour la gestion de la relation client*
- *Optimisation mix marketing produits,*
- *Optimisation de la Customer Experience*
- *Prévention des pannes, optimisation des processus de livraison*

### **Domaine Public**

- *Prévention criminalité*
- *Campagnes électorales*
- *Gestion des politiques publiques territoriales*
- *Ou sectorielles (climat)*
- *Open Data*
- *Pilotage de la consommation*

En voici quelques exemples d'application :

### **Analyse de campagnes marketing**

Un marketeur identifiera et touchera d'autant plus de cibles "granulaires" qu'il disposera d'informations. Le Big Data peut être utilisé pour analyser d'énormes quantités de données qui échappent aux solutions relationnelles classiques, de telle sorte que les spécialistes dans le domaine du marketing peuvent maintenant mieux repérer un public cible et associer les produits et services appropriés à un individu précis. Grâce au Big Data, il est possible d'étudier d'importants volumes d'informations à partir de nouvelles sources, comme le parcours de navigation ou les enregistrements des détails des appels, ce qui leur permet de mieux comprendre les tendances et les comportements d'achat des consommateurs.

### **Analyse de la fidélisation et de la perte de clientèle**

Une augmentation du nombre de produits par client équivaut souvent à une diminution de la perte de clientèle, et de nombreuses sociétés entreprennent d'importants efforts pour améliorer les techniques de vente croisée et de montée en gamme. Toutefois, l'étude de la clientèle et des produits à travers les secteurs d'activité s'avère souvent difficile, puisque des formats hétérogènes de données et des problématiques de gouvernance limitent ces efforts. Certaines entreprises ont la possibilité de charger ces données dans un cluster Hadoop, afin d'effectuer des analyses à grande échelle pour identifier les tendances. Le résultat montre les clients susceptibles de partir à la concurrence ou, encore mieux, ceux qui vont probablement approfondir leur relation commerciale avec l'entreprise. Des mesures peuvent alors être adoptées pour reconquérir ou encourager ces clients selon le cas.

### **Analyse des graphes sociaux**

Chaque réseau social ou communauté compte des utilisateurs ordinaires et des super utilisateurs, et reconnaître ces derniers est une tâche difficile. Avec le Big Data, les données issues des activités des réseaux sociaux sont explorées pour indiquer les membres exerçant une influence majeure sur le groupe. Ceci permet aux entreprises d'identifier les clients « les plus significatifs », qui ne sont pas forcément ceux utilisant l'offre de produits la plus vaste ou bénéficiant du budget conséquent, contrairement à la définition classique répandue dans le cadre de l'analyse décisionnelle.

### **Analyse des marchés des capitaux**

Que nous recherchions de grands indicateurs économiques, ou des indicateurs de marché spécifiques ou bien encore des avis sur une entreprise donnée et ses actions, la richesse des informations à analyser est impressionnante tant en provenance des sources classiques que des nouveaux réseaux. Si l'analyse par mots clé de base et l'extraction d'entités sont utilisées depuis plusieurs années, l'association d'informations classiques et de sources inédites telles que Twitter et d'autres médias sociaux permettent d'accéder à un aperçu détaillé de l'opinion publique, pratiquement en temps réel. Aujourd'hui, la plupart des institutions financières se servent, à différents degrés, de l'analyse des sentiments pour mesurer la perception du public sur leur entreprise, sur le marché, ou sur l'économie en général.

### **Analyse prédictive**

Afin de prévoir les changements économiques, les experts dans le domaine des marchés des capitaux confrontent d'un côté les algorithmes de corrélation avancés et calculs des probabilités, et, de l'autre, les données historiques et actuelles. Le volume important des archives d'informations sur les marchés ainsi que la vitesse exigée pour l'évaluation des nouveaux enseignements (par exemple : valorisations complexes d'instruments dérivés) font de l'analyse prédictive un problème majeur que le Big Data contribue à résoudre. En effet, grâce à la capacité à effectuer ce type de calculs plus rapidement, et avec du matériel informatique courant, le Big Data remplace de manière fiable l'approche relativement lente et coûteuse fondée sur les systèmes traditionnels.

### **Gestion des risques**

Les entreprises dont la technologie se veut être avancée et déterminée tentent de minimiser les menaces au moyen d'une gestion continue des risques et d'une analyse à large spectre des facteurs de risque, en croisant de vastes catégories de données. Par ailleurs, une demande de plus en plus pressante oblige à accélérer l'analyse des informations, malgré leur volume toujours croissant. Les technologies de Big Data s'imposent dans la résolution de ce problème : en effet, les calculs peuvent être effectués tout en accédant aux données. Qu'il s'agisse d'analyse croisée ou d'intégration d'informations sur les risques et les tendances financières, afin de rajuster les rendements et les bilans, il est nécessaire de fusionner, de permettre l'accès et de traiter à tout moment une quantité grandissante de données provenant des différents services autonomes de l'entreprise.

### **Trading**

Une analyse approfondie reliant les données comptables aux systèmes de repérage et de gestion des commandes peut fournir des informations stratégiques précieuses qui ne seraient pas disponibles avec les outils classiques. Afin de les identifier, une masse importante de données doit être traitée presque en temps réel à partir de sources multiples et hétérogènes. Cette fonction permettant de puissants calculs peut maintenant être effectuée par le biais des technologies Big Data.

### **Détection des fraudes**

Mettre en rapport des données à partir de sources multiples et non reliées augmente la possibilité d'identifier des activités frauduleuses. Si, dans le cadre du Big Data, l'on relie par exemple des mouvements bancaires effectués en ligne, aux distributeurs automatiques, via smartphone, par carte de paiement, à l'analyse du comportement web retracé sur le site de la banque où ailleurs, la détection des fraudes en est améliorée.

### **Services bancaires de détail**

Dans le domaine des services bancaires de détail, la capacité de déterminer avec précision le niveau de risque sur le profil d'un individu ou sur un prêt joue un rôle primordial dans la décision d'attribuer (ou de refuser) à un client certaines prestations. Comprendre correctement la situation protège non seulement la banque, mais satisfait aussi le client. Un accès à des informations exhaustives sur la clientèle permet aux banques de bénéficier de garanties et de visibilité afin de mieux cibler les offres de services. Il est aussi possible de prévoir les événements significatifs dans la vie du client, tel un mariage, la naissance d'un enfant, l'achat d'une maison, ce qui est un atout pour appuyer les activités de vente croisée et de montée en gamme.

### **Surveillance du réseau**

Tous les types de réseaux, qu'il s'agisse de transports, de communications ou de protection policière, peuvent bénéficier d'une meilleure analyse, activité dans laquelle interviennent les technologies Big Data. Considérons par exemple le réseau local d'un bureau : grâce au Big Data, des volumes considérables d'informations sont acheminés depuis des serveurs, des périphériques et du matériel informatique divers. Les administrateurs peuvent ainsi surveiller l'activité du réseau et détecter des congestions et bien d'autres problèmes avant qu'ils n'aient un impact négatif sur la productivité.

### **Recherche et développement**

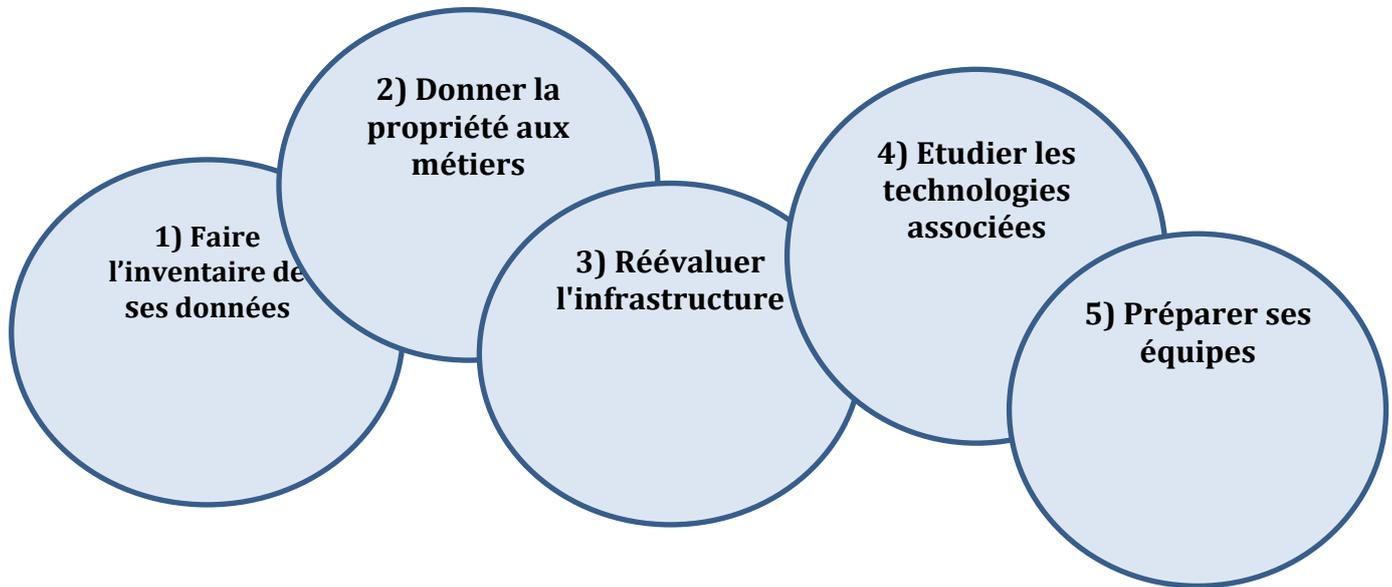
Les entreprises qui disposent de services de recherche et développement importants, comme les établissements pharmaceutiques, se servent des technologies Big Data pour examiner minutieusement d'énormes quantités d'informations texte et de données historiques afin d'accompagner la conception de nouveaux produits.

### **Industrie pharmaceutique**

Réduction et maîtrise des dépenses de santé, pertinence des prescriptions médicales, sécurité des patients... Le secteur de la pharmacie est au cœur de tous ces enjeux contemporains.

Les gigantesques flux de données liés aux prescriptions et aux diagnostics, et les multiples combinaisons sur les pathologies pourraient produire des informations capitales. Non seulement ces données permettraient de mieux connaître les effets des médicaments et d'obtenir une photo des populations, mais elles montreraient aussi les voies les plus intéressantes à explorer pour la recherche.

## Cinq actions pour mettre en place une démarche Big Data



### **Faire l'inventaire de ses données**

Toutes les entreprises, ou à peu près, ont accès à un flux régulier de données non structurées, que celles-ci proviennent des médias sociaux, ou de capteurs surveillant une usine, par exemple. Cela ne signifie pas qu'elles doivent toutes les sauvegarder et les exploiter. Cet afflux soudain a suscité un besoin artificiel d'essayer de comprendre toutes ces données, remarque Neil Raden, un analyste de Constellation Research. Ce souci a pu être provoqué par des consultants ou fournisseurs soucieux de promouvoir leurs solutions. « Il y a une certaine pression de la part de ceux qui commercialisent la technologie », observe Neil Raden. *Une première étape pourrait donc consister à inventorier quelles données sont créées en interne et déterminer quelles sont les sources externes, s'il y en a, qui pourraient apporter des informations complémentaires sur l'activité de l'entreprise.*

Une fois cet état des lieux engagé, les équipes informatiques devraient entreprendre des projets très ciblés qui pourraient être utilisés pour montrer quels résultats on peut obtenir, plutôt que d'opter pour des projets en mode big-bang.

### **Donner la priorité aux métiers**

Aligner l'IT avec les métiers est indispensable dans le cas d'un chantier aussi important que peut l'être le traitement des Big Data. Les premières occasions de traiter ces volumes se sont trouvées hors du périmètre de la IT. Par exemple dans des départements marketing qui ont décidé de récolter les flux des médias sociaux pour gagner en visibilité sur les besoins des clients et les tendances en matière d'achats. Si c'est effectivement du côté business que l'on devrait identifier l'intérêt de ces analyses, c'est à la IT que revient de prendre en charge la fédération et le partage des données et de mettre en œuvre la stratégie Big Data. *En même temps, ce n'est pas une démarche que la IT peut faire seule de son côté. « Il sera difficile d'en faire une success story si le projet n'est pas aligné sur les objectifs business ».*

## Réévaluer l'infrastructure

Dans la plupart des entreprises, les projets Big Data vont demander des changements importants. D'une part sur les infrastructures serveurs et de stockage, d'autre part sur la gestion des données, si l'on en croit Mark Beyer, du cabinet Gartner, ainsi que d'autres experts. Les responsables informatiques doivent se préparer à étendre leurs systèmes pour qu'ils supportent des bases en perpétuelle expansion, recevant données structurées et non structurées. Cela signifie qu'il faut trouver la meilleure approche pour rendre les systèmes à la fois extensibles et évolutifs et qu'il faut élaborer une feuille de route pour intégrer tous les systèmes disparates qui vont alimenter l'effort d'analyse des Big Data.

« Actuellement, les entreprises ont souvent des systèmes disparates et séparés pour gérer la paie, les relations clients, le marketing », indique Anjul Bhambhri, vice-président, responsable des produits Big Data chez IBM. Les CIO ont besoin de mettre en place une stratégie pour les faire converger. « Il faut pouvoir poser des questions qui traversent tous ces systèmes pour obtenir des réponses », précise le VP d'IBM.

## Etudier les technologies associées

Evidemment, le monde des Big Data apporte sa liste d'acronymes et de technologies. Les outils en Open Source sont ceux qui retiennent le plus l'attention. On met au crédit des Hadoop, MapReduce et NoSQL d'avoir aidé les géants du web que sont Google et Facebook à naviguer à travers leurs réservoirs de Big Data. La plupart de ces technologies, même si elles sont désormais disponibles sous une forme commerciale, sont encore assez immatures et requièrent pour s'en servir des compétences très spécifiques. Parmi les autres technologies importantes dans le monde des Big Data figurent l'analytique in-database pour être au plus près des données, les bases exploitant le stockage en colonnes et les appliances de datawarehouse. Les équipes IT vont devoir comprendre ces nouveaux outils pour pouvoir faire des choix avisés en matière de Big Data.

## Préparer ses équipes

Qu'elles aient besoin d'experts Hadoop ou de spécialistes des données (data scientists), les départements IT manquent sévèrement des talents nécessaires pour aller plus loin avec les Big Data. Les compétences en matière d'analytique sont peut-être les plus cruciales et c'est là que se trouvent les manques les plus importants.

*McKinsey prévoit que, pour les seuls Etats-Unis, on aura besoin d'ici 2018 de 140 000 à 190 000 experts supplémentaires spécialisés dans les méthodes statistiques et l'analyse de données.* Parmi les postes les plus demandés se trouvera la fonction de plus en plus médiatisée de « data scientist ». En outre, McKinsey s'attend à des demandes à la fois du côté métier et technique pour quelque 1,5 million de managers à l'aise avec les données, disposant d'une formation reconnue dans l'analyse prédictive et les statistiques.

Les directions des départements IT vont devoir elles-mêmes se transformer pour exceller dans ce nouveau monde. Les managers IT du futur combineront des compétences sur l'analyse de données et les processus métiers, estime Mark Beyer, de Gartner. « Les CIO ont eu l'habitude de gérer des infrastructures à partir des prescriptions des métiers, par opposition à un CIO qui serait capable d'identifier une opportunité et par conséquent de pousser vers une utilisation innovante de l'information », explique-t-il. C'est la transformation qui devra très probablement se produire.

## Les compétences nécessaires pour une démarche Big Data<sup>2</sup>

### **Data scientists**

Les data scientists constituent une catégorie de professionnels capables de mettre en oeuvre leurs techniques d'organisation de données dans des applications métiers. Ils doivent aussi savoir communiquer pour expliquer les résultats trouvés dans les données à la fois à la DSI et aux directions métiers. Typiquement, ces data scientists ont leur propre sandbox dans laquelle ils explorent et examinent les données de l'entreprise en accompagnant la démarche d'innovation. Anjul Bhambhri, vice-président responsable des produits Big Data chez IBM, les décrit de façon imagée : « en partie analyste et en partie artiste, un data scientist a de la curiosité, il observe les données et repère des tendances ». C'est, selon lui, un profil qui veut vraiment apprendre et transformer une entreprise.

### **Spécialistes des outils d'ETL**

Avec l'explosion de données de toutes sortes dont les entreprises cherchent à tirer parti, le besoin en professionnels capables de récupérer et d'intégrer ces Big Data a augmenté de façon significative. Parmi eux figurent notamment les familiers des solutions d'ETL, ces outils permettant d'extraire les données de leur source initiale, de les mettre au bon format et de les charger dans la base cible (extraction, transformation, loading). Des logiciels tels qu'en proposent le Français Talend, ou encore Informatica et Pentaho pour ne citer qu'eux. Comme les logiciels d'ETL ont acquis de la maturité, les profils spécialisés sur ces outils sont des compétences dont on aura besoin dans la durée.

### **Développeurs Hadoop**

Le mouvement Big Data a enregistré une hausse de la demande autour du framework Open Source Hadoop, adapté au traitement des jeux de données très volumineux, et de toutes les technologies associées comme Hive, HBase, MapReduce, Pig, etc. Le traitement de ces teraoctets ou petaoctets serait trop coûteux avec des outils de BI classiques et prendrait beaucoup trop de temps si on ne recourait pas aux capacités massivement distribuées d'Hadoop. Dans l'actuel paysage Big Data, les personnes possédant une expérience sur ce framework disponible dans de nombreuses distributions (dont Hortonworks et Cloudera) sont les plus recherchées, selon Greg Jones.

### **Spécialistes des outils de visualisation**

Analyser d'énormes quantités de données s'avère souvent difficile. Les nouveaux outils de visualisation tels qu'en proposent Tibco Spotfire, Qlickview ou Tableau Software permettent une exploration rapide et intuitive de ces données. On peut penser que les compétences requises pour les utiliser se rapprochent de celles d'un profil spécialisé en BI. Mais, précise KForce, avec la montée en force d'Hadoop, cela représente une nouvelle catégorie de profils spécialisés. Pour l'instant, ces ressources sont recherchées à court terme, en externe. Avec le temps et la maturité des outils, la demande pour ces profils se modérera et ces ressources seront plutôt utilisées en interne, estime Greg Jones.

### **Développeurs OLAP**

L'expertise des développeurs OLAP consiste à optimiser l'organisation des données. Ce sont eux qui récupèrent les données de sources relationnelles ou non structurées pour créer des modèles multidimensionnels -souvent désignés sous le nom de schémas en étoile ou en flocon de neige- et qui construisent ensuite l'interface utilisateur pour accéder aux données à partir de requêtes prédéfinies.

---

<sup>2</sup> La valorisation des données de l'entreprise ne saurait se réduire au simple enjeu technologique du Big Data. Pour l'éviter, elle doit être confiée à des Data Scientists capables d'appréhender la problématique dans toutes ses dimensions : métier, informatique, statistique et mathématique.

## Pourquoi le big data ne tient pas encore ses promesses

Aujourd'hui une dizaine d'entreprises ont déjà lancé une démarche Big Data, cependant, elles sont encore loin de profiter d'un tel phénomène aussi prometteur.

En effet, les technologies sont là, pas les hommes. Matteo Pacca, du cabinet McKinsey, est l'un des plus fervents partisans du Big Data. Depuis deux ans, il sillonne les conférences internationales pour vanter les bénéfices que tireront les entreprises du traitement massif de données informatiques et démontrer, prévisions à l'appui, qu'il deviendra pour elles le moyen ultime de se différencier. Pourtant, pour la première fois, en ouverture du dernier salon parisien consacré au Big Data, son discours si volontaire laissait planer comme un léger doute. Du bout des lèvres, il s'est rendu à l'évidence : les acteurs ne savent pas encore tirer parti du concept. Ils ont même du mal à dire en quoi celui-ci pourrait les aider. *“Un cabinet de recrutement m'a contacté car il recherchait des profils de responsables de données (Chief Data Officer – NDLR) pour l'un de ses clients désireux de se lancer, expliquait-il en ouverture du salon. Mais ni le cabinet ni la société concernée n'étaient en mesure de définir précisément le contour de ce poste.”*

Une anecdote qui résume l'énorme décalage entre l'incompréhension des entreprises envers ce qu'est le Big Data et l'ultramédiatisation d'un phénomène qui fait la une des plus grands magazines américains (dont la prestigieuse Harvard Business Review).

Le Big Data affiche des prévisions de croissance de 15 à 20 % par an, à faire pâlir les autres segments du secteur numérique. Ne pas y aller, c'est se priver d'un fantastique outil. Mais comment faire ? Il existe au moins cinq raisons pour lesquelles les organisations rechignent à se lancer dans l'aventure.

### **Les compétences nécessaires sont encore rares**

Savoir tirer parti d'un nombre considérable d'informations disparates exige une triple sensibilité : il est indispensable d'aimer la statistique et l'informatique, et de connaître les besoins de la personne à laquelle on s'adresse (responsable marketing, commercial...). Autant le dire, ces profils polyvalents (baptisés Data Scientists) sont très recherchés. Leur rôle : déterminer les variables qui comptent, se poser les bonnes questions et repérer les tendances anormales (fraudes, dysfonctionnement de site Web, etc.).

Le tout en maîtrisant les nouvelles techniques de programmation indispensables pour extraire les informations pertinentes... *“ Les quelques profils couvrant l'étendue du spectre émargent entre 1 500 et 1 800 euros par jour, contre 1 200 pour un prestataire classique. Rares sont les entreprises prêtes à payer ce prix. Et lorsqu'elles les font intervenir, c'est seulement pour une opération commando ”*, explique Jeremy Harroch, fondateur de Quantmetry, spécialisé dans l'analyse de données.

### **Les responsables opérationnels restent difficiles à convaincre**

Les rares projets français de Big Data émanent soit d'une direction informatique, soit d'une cellule innovation. Dans tous les cas, il s'agit de techniciens qui ont d'abord lancé une initiative avant de tenter de convertir leurs directions métier. Quitte à s'arracher parfois les cheveux. *“Pour les responsables opérationnels auxquels nous avons proposé de nouveaux services liés à la géolocalisation, nous faisons figure de poil à gratter, sourit Mathieu Gras, en charge du Big Data chez SFR. Nous avions du mal à expliquer notre démarche, car c'était à eux d'imaginer les applications concrètes...”*

De la même façon, chez Bouygues Telecom, c'est d'abord la DSI qui a déployé la technologie de traitement de données, Hadoop. Elle souhaitait mesurer la qualité de services de ses sites Web et comprendre comment améliorer les temps de réponse des pages. Son expérimentation a rapidement séduit d'autres branches de l'organisation, qui ont profité de l'aubaine. Le service client s'appuie

aujourd'hui sur la même plate-forme et les mêmes données pour optimiser le parcours des internautes sur les différents sites du groupe.

### **Toutes les données ne sont pas exploitables**

Si les données existent en masse, rares sont celles directement exploitables en l'état. Et c'est un vrai problème, car il est alors nécessaire de les retraiter, ce qui engendre des surcoûts souvent négligés lors de la définition du projet. *“ Les données tirées des réseaux sociaux ou celles reflétant l'activité des internautes sur les pages Web (on parle alors de logs) demandent à être un minimum nettoyées ”*, précise Florian Douetteau. Certes, ce sujet de la qualité de la donnée n'est pas nouveau. *“ Ce qui l'est, c'est le volume à traiter dans des temps toujours plus courts ”*, poursuit le créateur de Dataiku.

Autre chantier souvent sous-estimé : parvenir à mettre en relation des informations piochées à différents endroits. *“ C'est très bien de les stocker en masse, lance Serge Boulet, responsable marketing chez l'éditeur SAS. Mais comment, par exemple, rapprocher autour d'un même client des données résultant d'une navigation Web, d'une géolocalisation ou d'un ticket d'appel ? L'opération n'a rien d'évident. ”*

### **La rentabilité des projets n'est pas toujours au rendez-vous**

*La question du retour sur investissement semble presque incongrue pour les partisans du Big Data, dont la devise pourrait être : “ Investissez d'abord, vous trouverez bien à l'utiliser ensuite. ” En effet, seule l'exploration des données disponibles permet de déduire a posteriori les questions intéressantes et les futures applications. En temps de crise, un tel message a de quoi rebuter les directions générales. “ Aucune solution miracle ne garantit la rentabilité d'un projet de Big Data, concède Frédéric Brajon, associé chez CGI Business Consulting (ex-Logica). Il faut tester les idées dès qu'elles apparaissent pour identifier très vite celles qui méritent d'être creusées. C'est un changement de méthodologie ”.*

### **L'offre technique se révèle pléthorique et complexe**

L'expression Big Data est à la mode. Tous les acteurs du secteur technologique veulent avoir leur offre. Résultat : le paysage est très confus. Et cette profusion effraie les sociétés. *“ Il existe six à sept types de plates-formes d'exploitation des données. Et pour chacune de ces familles, comptez autant de fournisseurs ”*, résume Florian Douetteau. Pire, chacun de ces outils correspond à des besoins spécifiques. *“ Les uns permettent d'aller vite, les autres de traiter de gros volumes, les troisièmes des données très disparates... Aucune solution du marché ne sait tout faire correctement ”*, reconnaît Michel Bruley, directeur marketing Europe chez Teradata.

Pour ajouter à la confusion, tous les fournisseurs, même ceux qui ne proposent pas de nouvelles technologies, ont inondé le marché avec des messages marketing ciblés sur le Big Data. Jusqu'à l'écœurement... *“ Bien des acteurs ont profité de cette vague pour repackager des solutions des années 70 ! Ce qui a, évidemment, créé une désillusion proportionnelle à l'engouement suscité ”*, déplore Jean-Baptiste Dezard, responsable marketing logiciel d'IBM.

## Focus, Big Data et le secteur bancaire

---

Bien que le métier du banquier reste par nature inchangé (assurer l'intermédiation entre ressources et besoins de financement), les solutions Big Data ouvrent de nouvelles perspectives. Nous décrivons dans ce qui suit les opportunités permises par le Big Data.

### **Introduction**

L'évolution des Directions des Risques des grandes industries (banques et assurances notamment) doit être mise en parallèle avec celles des données et des IT, et sera conditionnée par la capacité d'autres fonctions à intégrer ces changements (directions commerciales et marketing notamment).

#### Evolution des Directions des Risques

Meilleure intégration dans l'organisation globale de la Banque

Meilleure structuration de « l'offre de la filière Risque » auprès des Métiers (*comment et sur quoi cette dernière se propose d'intervenir auprès des métiers, à leur demande*)

Plus de valorisation des résultats obtenus grâce à des indicateurs ajustés et dynamiques

#### Evolution des capacités IT des Institutions

Mutualisation de l'architecture SI dans la plupart des banques (plus homogène)

Marché de solutions clés en main avec simplification de la mise en œuvre, notamment pour la mise en conformité aux attentes réglementaires

Puissance de calculs et préservation des données exponentielles

#### Evolution de l'approche aux données

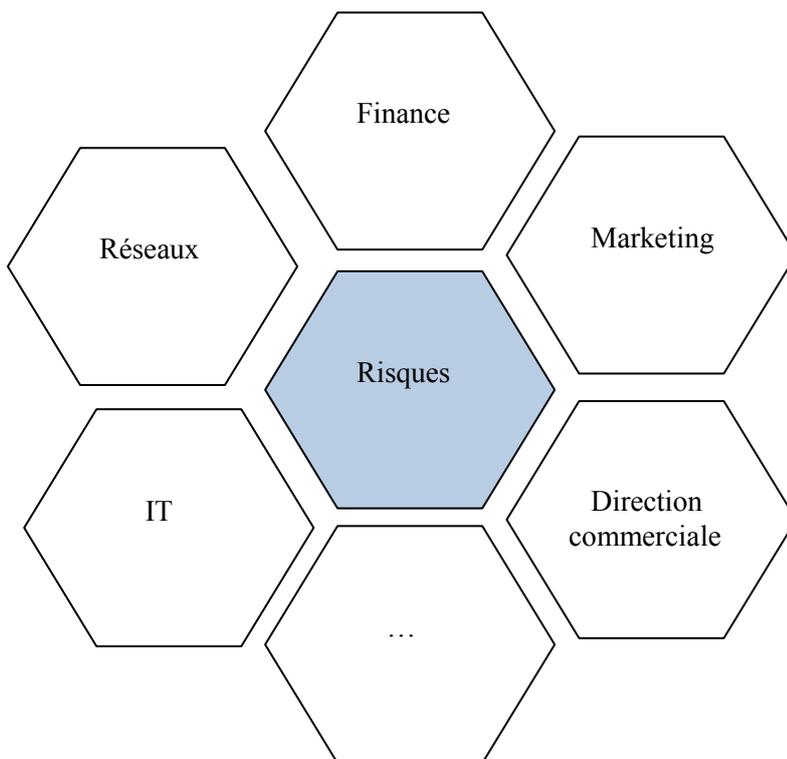
Amélioration constante de la qualité des données

Stockage transversal des données

Vers une exploitation des données comportementales (réseaux sociaux, etc.)

**Big Data, une fonction Risques non isolée**

Le Big Data prévoit de mettre un lien fort entre la direction des Risques et les autres directions de l'établissement financier.



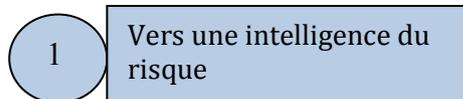
De là, on tire les points suivants :

- Lien marketing – risques plus fort grâce aux grands projets informatique CRM
- Un marketing plus ciblé grâce à une meilleure segmentation des clients
- Data Warehouse unique pour toutes les fonctions de la banque
- Après la crise durable, nécessité de sensibiliser largement aux enjeux de liquidité
- Compréhension fine des risques ALM, de marché, de crédit par la fonction Finance
- Multiplication des projets communs (ex : gestion du LCR)
- Mesure de la rentabilité du réseau d'agences grâce à la collecte de data clients
- Meilleure gestion du risque opérationnel réseau
- ...

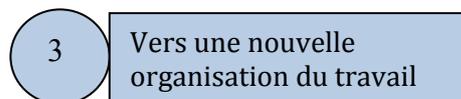
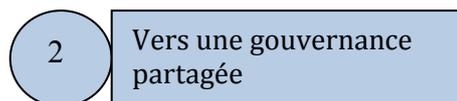
### **Big Data, une fonction Risques organisée et intelligente**

Les risques sont dorénavant placés au service de la stratégie de la banque en lien direct avec les organes décisionnels et les autres directions (via des Comités partagés), ce qui crée une nouvelle intelligence du risque.

Des fonctions régaliennes qui vont rester essentielles...



... mais dont les modalités vont radicalement changer

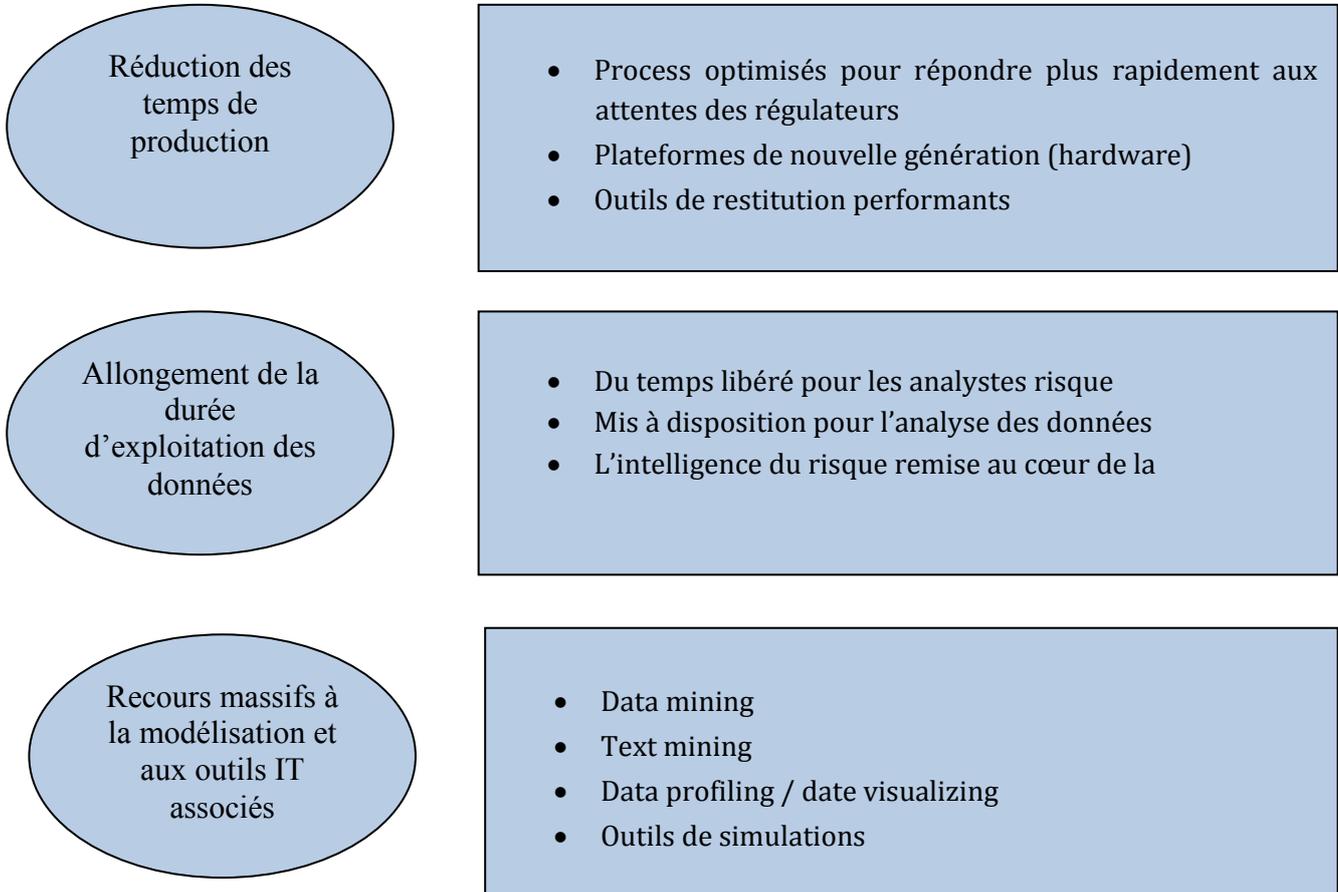


Il s'en sort :

- Naissance / développement de départements d'études => analyse du risque, traitement et analyse de la donnée
- Une réduction des process => Réduction des temps de production versus un allongement de la durée d'exploitation et d'analyse des données
- Importance de la modélisation (data mining, profiling, simulation ...)
- Mise en place de relais beaucoup plus forts avec les autres directions de la banque (Comités partagés et/ou multilatéraux, études conjointes et reportings partagés ...)
- Diminution des échelles hiérarchiques

**Big Data, des process de production simples**

Cette simplification attendue des process se fera en contrepartie d'un allongement des étapes à valeur ajoutée de ces mêmes process.

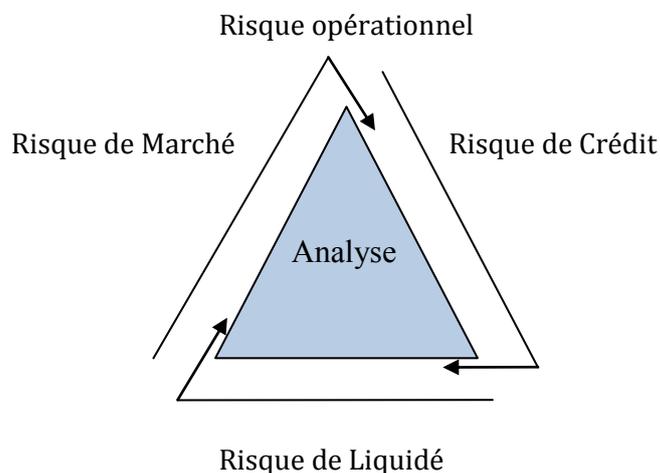
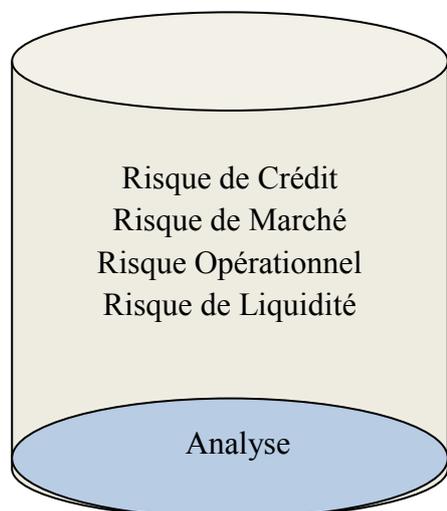


### Anticipation sur l'évolution d'une fonction Risques (en environnement Big Data)

- D'une analyse compartimentée vers une analyse transversale des risques

D'une analyse compartimentée où chaque famille de risques est analysée séparément et indépendamment ...

... vers une analyse transversale où l'on est capable d'identifier et d'appréhender de façon globale tous les risques à la fois



- Un nouveau format de suivi et de pilotage du risque

Formats de suivi intelligents intégrant une vision plus consolidée

Grâce aux nouvelles expertises et compétences de la Direction des Risques  
*Des reporting moins segmentés mais aussi plus précis et concis capables de mettre en exergue les risques à un moment donné*

Demande pour des analyses d'anticipation et de prédiction

Sur des horizons aussi bien court terme que stratégiques  
Mise en perspective des risques avec d'autres dimensions  
*Rentabilité, Marge opérationnelle, capacité commerciale, etc.*

Nouvelles limites, nouveaux indicateurs et ratios de surveillance

Moins axés sur les aspects réglementaires, plus flexibles  
Evolutifs (non figés) et donc adaptés au niveau général de risque de la banque à un moment donné (ex : liquidité)  
Liés à l'activité réelle de la banque et sa prise de risque (ex : transformation)

Nouveau format de reporting

Une plus forte capacité de synthèse grâce au tri de l'information  
Un reporting dynamique et interactif  
*Importance du détail et de la profondeur de la donnée*  
*Exemple : zoom sur une donnée particulière pour comprendre l'augmentation des RWA sur une activité de marché donnée*

- Vers des analyses plus avancées

FORMAT DE  
RESTITUTION  
« INTELLIGENT »

- Embrasser tous les risques de l'établissement en « ½ graphs »
- Représentation visuelle intelligente
- Format 3D et 4D (intégration du mouvement => projections)

REPORTING A  
POSSIBILITES  
CALCULATOIRES

- Formules intégrés dans un graph (simulation des effets)
- Interactions entre les différents indicateurs et mise à jour instantanée (changement d'un scénario=> recalcul immédiat des niveaux de RWA, liquidité sur tous les tableaux)
- Caractère auto correctif : Détection automatique des incohérences, écarts entre indicateurs, ... (caractère auto correctif)

FORMAT DE  
REPORTING  
« FLOTTANT »  
(non figés)

- Indicateurs automatiquement choisis en fonction de l'évolution des risques
- Tableaux de bord interactifs (changements d'axes en temps réel)
- Détection des zones de risques et fixation automatique des limites adéquates (auto détection du risque...)
- Construction de reporting par objectif pré-fixés (reporting stratégique, reporting détection du risque, reporting mesure du risque)

- Vers l'utilisation des techniques statistiques avancées

Scoring intégrant les comportements sociaux

- Adaptation dynamique des outils de détection de la fraude
- Capacité d'adaptation des process de recouvrement beaucoup plus souples et rapides (sélection des garanties, etc.)
- Récupération automatique des données exogènes à l'établissement et utiles pour son activité

Développement de modèles complexes et plus prédictifs

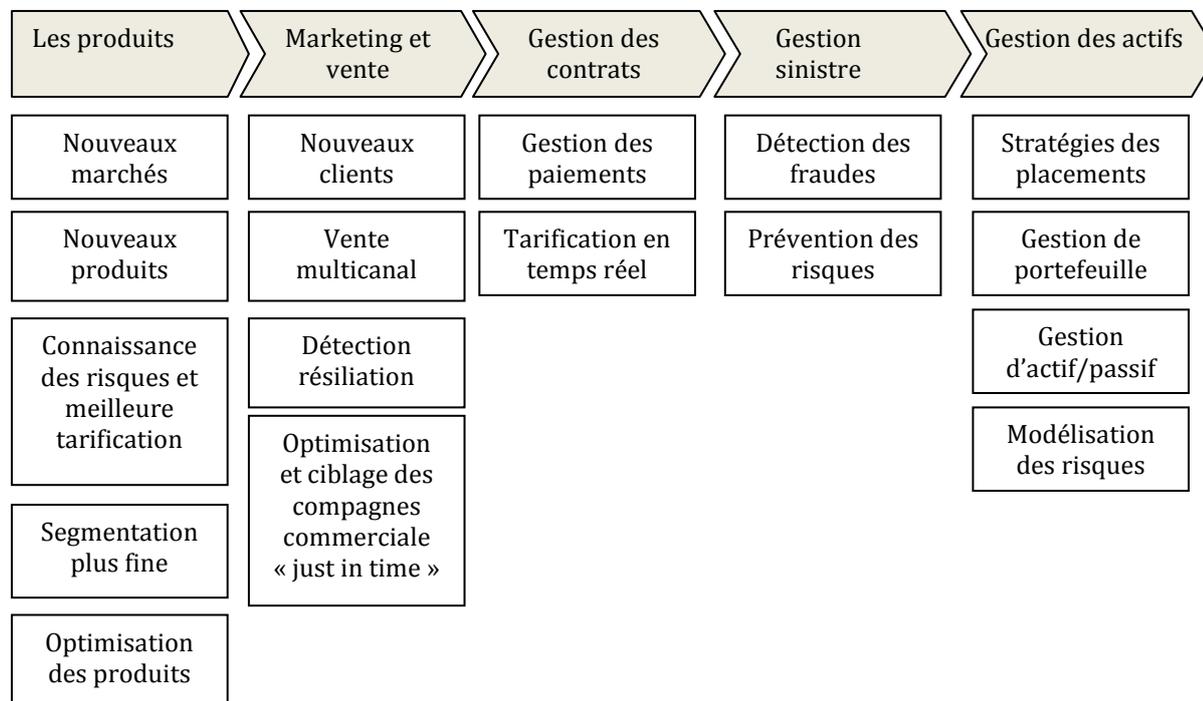
- Réseaux neuronaux (auto apprentissage => exemple de la fraude) + recherches des causalités dans les données accessibles
- Data mining à grande échelle
- Simulation (Monte Carlo ou équivalents) sur toutes les données du Datawarehouse
- Développement de nouvelles approches de modélisation comportementale (text mining)

Utilisation des outils de risque dans la gestion quotidienne

- Couverture du risque (provisions, ...)
- Réduction du risque : sélection à l'entrée
- Maîtrise de la rentabilité / risque : montant, totaux ou volume d'un prêt pour un client donné
- Simulation du risque : incidence en risque d'une campagne marketing
- ....

## Focus, Big Data et le secteur assurantiel

### Principales opportunités pour l'assurance :



### Zoom, Big Data dans la tarification en assurance :

#### Tarification prime pure et données

Etablissement prime pure : recherche de données permettant d'expliquer la sinistralité grâce au :

- GLM
- Classification pour zonier et véhiculier

Données traditionnellement utilisées pour l'établissement de la prime pure limitées :

- Données essentiellement internes
- Faible recours à des données externes

Mouvement vers l'utilisation de plus données :

- Variables croisées : âge x sexe
- Interne : croisement de données entre produits :
  - Auto / MRH : âge x propriétaire/location résidence principale : variable tarifaire en auto
  - Données bancaires / prêt et données assurances
- Externe :
  - INSEE : densité, éloignement des services essentiels...
  - Données véhicules : SIV

Existe-t-il des données BIG DATA permettant de mieux expliquer la sinistralité que les données existantes ?

Remise en cause du modèle utilisé aujourd'hui (GLM)...

## Positionnement tarifaire et données

Positionnement tarifaire : recherche de données permettant d'expliquer le choix de l'assuré :

- Modèle d'élasticité au prix de la demande
- Positionnement concurrentiel
- Optimisation tarifaire

Ces modèles très utilisés au Royaume-Uni, sont encore très peu utilisés en France, néanmoins cela devrait évoluer :

- Loi Hamon -> potentiel effet ciseau (Comment fidéliser le portefeuille ?)
- Comparateurs -> potentiel anti-sélection (Comment acquérir les profils à valeur ajoutée ?)

Ces modèles sont très gourmands en données :

- Données connaissance du risque
- Données relation client
- Données flux primes / sinistres avec effet retard

Existe-t-il des données BIG DATA permettant de mieux expliquer la propension des assurés à souscrire / renouveler ?

## Quel apport de données externes ?

L'utilisation de données externes (sans parler de BIG DATA) est actuellement faible et limitée :

- Données INSEE, données délinquance (CartoCrime) ...
- Données Véhicules : constructeurs, ...

Ces données permettent d'acquérir des variables complémentaires qui permettent de qualifier plus finement le risque qu'à partir des données transmises par le réseau et/ou l'assuré :

- La ville de résidence est-elle en zone rurale peu peuplée ou en zone urbaine ? Est-ce un lieu où la criminalité est faible ou élevée ?
- Le véhicule est-il une :
  - citadine : pour aller faire les courses le week-end ?
  - sportive : pour épater les copains et sortir en boîte ?
  - familiale : pour partir en vacances avec la famille ?
  - commerciale : pour faire des tournées en province ?

Tout apport d'information permettant de renforcer la caractérisation d'un profil de risque est utile pour éviter l'anti-sélection

Parmi aussi les données BIG DATA, il existe une catégorie qui regroupe les qualités recherchées permettant de mieux caractériser le risque. Plutôt que de se baser sur la recherche de variable permettant de traduire un comportement à risque, cette catégorie cherche à capturer le comportement à risque directement à la source Les capteurs :

- Auto : OBD-II (comportement de conduite : freinage brusque, coups de volant intempestifs)
- MRH (dommage électrique d'euro-assurance): Domotique (comportement des occupants)
- Santé : capteur de données physiologiques (activité physique, intellectuelle...)

Les réseaux sociaux peuvent apporter beaucoup également concernant le positionnement tarifaire :

- Sensibilité, attentes (écoresponsable, développement durable...)
- Réceptivité aux messages marketing
- Réactivité face aux changements
- Fidélité
- Elasticité au prix...

## Big Data et le cadre réglementaire

Quelle protection des données à caractère personnel (permettant d'identifier une personne) ?

- Obligation déclarative auprès de la CNIL (données et usage),
  - Obligation d'information aux personnes (accès / suppression / rectification),
  - Obligation de sécurité (inventaire et information sur les violations de sécurité de données à caractère personnel).
- Attention : interdiction de principe sauf exceptions de transférer des données personnelles hors de l'UE => préciser localisation des serveurs dans les contrats

Quelle propriété ?

- Absence de propriété d'une donnée isolée (sauf protection contractuelle spécifique),
- Le producteur de la base dispose d'un droit de propriété qui résulte des investissements réalisés pour la constituer et qui lui permet d'interdire l'extraction ou la réutilisation,
- La structure de la base peut être protégée par le droit d'auteur si les conditions du droit d'auteur s'appliquent.

Quelle exploitation ?

- Attention : la valorisation des données suppose une collecte et un traitement licite de ces données,
- L'exploitant propriétaire de ses données peut librement les utiliser dans la limite de la protection des données personnelles,
- Dans le cas contraire, l'exploitation des données est limitée par les droits contractuellement conférés,
- Certaines exploitations des données sont elles-mêmes contraintes par la législation, en particulier le droit à la consommation et le droit à la concurrence s'appliquent en matière de personnalisation des offres commerciales.

## Conclusion

---

Le Big Data est une vraie promesse de valeur tractée par l'exploration et l'exploitation des cas d'usage métier au carrefour des données internes à l'entreprise et des données externes (Open data, Web Analytics...). Il englobe de nouvelles méthodes de travail, de nouvelles compétences, un effet décloisonnant dans les organisations qui vont devoir s'appuyer sur de nouvelles architectures informatiques. Sa puissance se résume dans :

- Analyser des données issues de différentes sources depuis une seule et même application;
- Explorer naturellement les associations entre les données;
- Visualiser les données à l'aide de graphiques soignés et performants;
- Accéder à vos données à partir de périphériques mobiles pour les analyser;
- Développer la prise de décisions collaboratives en temps réel et de façon sécurisée

**Nexialog Consulting**

Pour plus d'information sur le sujet,

<http://files.meetup.com/2243521/A%20Guide%20to%20Big%20Data%20in%20Finance%20-%20Entrepreneurs%20Edition%20v1.0.pdf>  
<http://www.data-business.fr/big-data-definition-enjeux-etudes-cas/>  
<http://www.data-business.fr/exemples-applications-impact-big-data-entreprise/>  
[http://www.lesechos.fr/opinions/points\\_vue/0203380497218-le-big-data-et-les-atouts-francais-658423.php](http://www.lesechos.fr/opinions/points_vue/0203380497218-le-big-data-et-les-atouts-francais-658423.php)  
<http://tempsreel.nouvelobs.com/economie/20140417.OBS4360/pourquoi-la-big-data-devient-un-tres-gros-business.html>  
<http://visionarymarketing.fr/blog/2014/04/big-data-marketing/>  
<http://www.soft-concept.com/surveymag/big-data-fin-des-etudes-marketing-4.htm>  
<http://www.altic.org/decouvrir-notre-offre/business-intelligence/big-analytics/425-big-data-pour-qui-pourquoi>  
<http://www.data-business.fr/big-data-analytics-les-facteurs-derriere-revolution-data/>  
<http://www.strategie.gouv.fr/blog/2013/11/note-analyse-des-big-data/>  
[http://info.talend.com/rs/talend/images/WP\\_FR\\_BD\\_Talend\\_4Pillars\\_BigDataManagement.pdf](http://info.talend.com/rs/talend/images/WP_FR_BD_Talend_4Pillars_BigDataManagement.pdf)  
<http://public.dhe.ibm.com/common/ssi/ecm/fr/gbe03519frfr/GBE03519FRFR.PDF>  
[http://www.data-business.fr/wp-content/Downloads/LivresBlancs/Guide\\_du\\_Big\\_Data\\_2013\\_2014.pdf](http://www.data-business.fr/wp-content/Downloads/LivresBlancs/Guide_du_Big_Data_2013_2014.pdf)  
[http://www.optimindwinter.com/wpcontent/themes/optimind/upload\\_dbem/2013/10/201310\\_Dossier\\_technique\\_Optimind\\_Winter\\_Big\\_Data.pdf](http://www.optimindwinter.com/wpcontent/themes/optimind/upload_dbem/2013/10/201310_Dossier_technique_Optimind_Winter_Big_Data.pdf)  
<http://www.revue-banque.fr/banque-detail-assurance/article/big-data-defis-opportunités-pour-les-assureurs>  
Big Data & Risques, Mise en perspective du Big Data dans la gestion des risques bancaires, Chappushalder & Cie.  
<http://fr.slideshare.net/StphaneChappellier/2014173127-big-dataetdonne769esexternesdanslesmode768lesdetarifcationvf>  
<http://www.altic.org/decouvrir-notre-offre/business-intelligence/big-analytics/118-hadoop/350-big-data-pour-vos-projets-decisionnels>  
<http://pro.01net.com/editorial/595059/pourquoi-le-big-data-ne-tient-pas-encore-ses-promesses/>