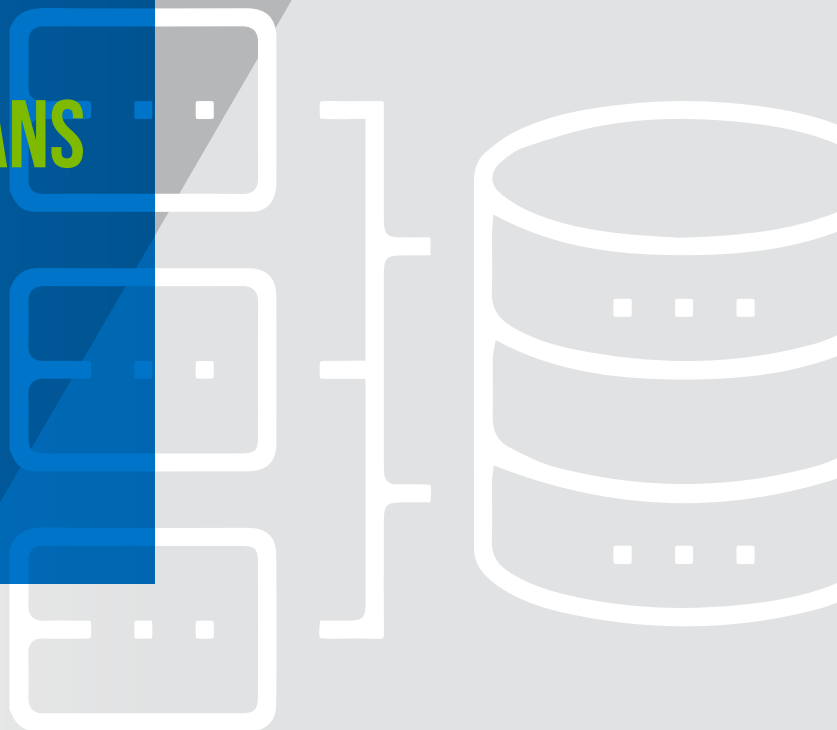


LIVRE BLANC

LE BIG DATA **EN ENTREPRISE :** **COMMENT SE LANCER DANS** **L'AVENTURE ET RÉUSSIR**

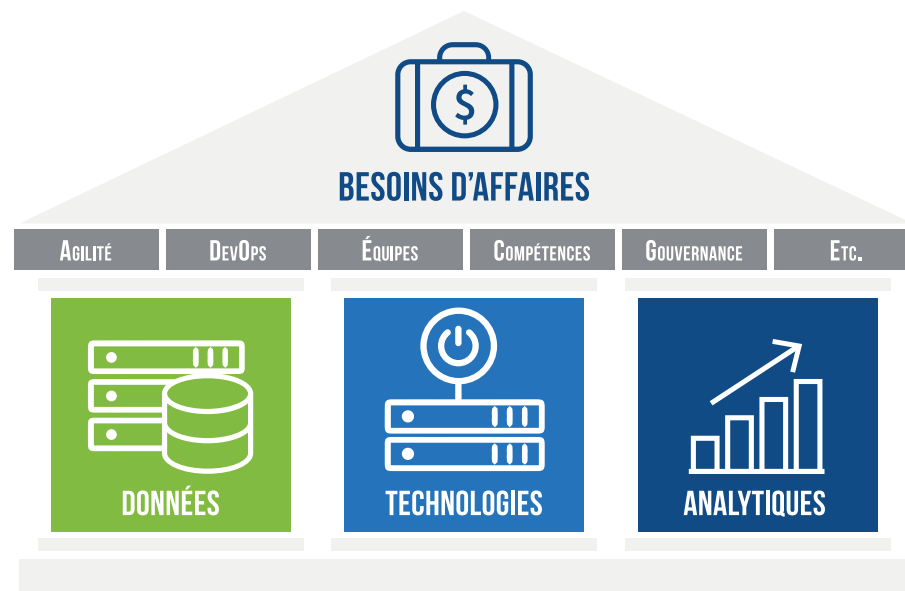
PAR : FRANÇOIS VIENNEAU BINETTE
Concepteur de solutions Big Data



NOVIPRO

TABLE DES MATIÈRES

- Introduction 4
- Qu’est-ce qu’un projet Big Data ? 5
- Pourquoi se lancer dans le Big Data ? 6
- L’impact sur l’entreprise 8
- La hausse des revenus 8
- La baisse des dépenses10
- Sans besoins d’affaires, oubliez le Big Data !11
- Un besoin d’affaires ?11
- Les besoins, ça part d’où ?12
- Déterminer ses besoins d’affaires13
- Une méthodologie de développement13
- Comprendre l’entreprise14
- Développer des idées14
- Synthétiser les besoins16
- Les cas d’usage de référence17
- Modernisation de l’architecture18
- Contrôle et compréhension19
- Optimisation et prédiction20
- Conclusion21



INTRODUCTION

« **Big Data** », voilà un mot qui fait grandement parler de lui depuis quelques années. À entendre certains, ce serait même le Saint Graal des entreprises. Bien que ce ne soit pas tout à fait faux, le Big Data (ou mégadonnées) demeure un domaine d'activité parfois incompris, que ce soit par sa nature à vouloir briser les silos de données des entreprises, ou simplement parce que les entreprises se demandent encore quoi faire de ces mégadonnées ! Mais c'est encore plus vrai depuis que l'on parle partout d'intelligence artificielle, d'apprentissage automatique et d'Internet des objets (IdO) ; des pratiques qui tirent leur origine de ces fameuses mégadonnées.

L'objectif de ce livre blanc est donc de présenter une méthodologie de développement d'idées et de besoins qui pourraient être comblés par des technologies Big Data. C'est-à-dire que nous présenterons une technique pour déterminer rapidement si ces nouvelles technologies peuvent être avantageuses pour l'entreprise qui s'intéresse au phénomène. Un des grands aspects de tout projet Big Data est la définition des besoins d'affaires « Qu'est-ce que je veux faire avec mes données ? », ce qui est souvent tenu pour acquis et négligé.

On ne peut présenter une méthodologie sans une mise en contexte. C'est pourquoi nous aborderons plusieurs sujets, notamment les raisons de se lancer dans un projet Big Data et ce à quoi un projet Big Data est censé répondre. Pour bien comprendre le contexte, il est pertinent de revoir les cas d'usage de référence, c'est-à-dire ce qui se fait dans les entreprises en matière de Big Data. Et finalement, nous allons élargir le sujet pour laisser place à l'innovation dans l'entreprise.

QU'EST-CE QU'UN PROJET BIG DATA ?

Avant d'aller plus loin, prenons le temps d'expliquer ce que l'on sous-entend par « projet Big Data ». Un projet, une initiative, une implantation Big Data, ou ce qu'on appelle « faire du Big Data », se fonde sur **trois piliers simples** et essentiels qui travaillent conjointement pour répondre à des **besoins d'affaires**.

Il s'agit des données qui sont bien sûr à la source de ce phénomène. Ce sont des données dont l'abondance, le format et la fréquence de production sont difficiles à maîtriser par les systèmes d'information (SI) traditionnels.¹

Il y a aussi les technologies qui permettent de capter, de stocker, de traiter et de transformer des volumes importants de données, et ce, de façon efficace, efficiente et au moment opportun. Il existe une grande variété de technologies dites Big Data sur le marché. L'évaluation se fait toujours en fonction de la facilité d'utilisation, de la convivialité, de la compatibilité et de l'intégration avec d'autres systèmes, des modèles de licences (ouvertes ou propriétaires), du support et de l'expertise disponibles, etc. Qui dit Big Data dit souvent technologies de type NoSQL conçues pour permettre le stockage de données de type semi ou non structurées.²

Il y a également les méthodes statistiques et algorithmiques permettant de transformer les données en information à valeur ajoutée. Souvent appelées Data Science dans le contexte Big Data, on retrouve ces méthodes sous d'autres techniques ou appellations comme l'apprentissage automatique (*Machine Learning*), l'intelligence artificielle, l'apprentissage en profondeur (*Deep Learning*), la visualisation de données (*Data Visualisation*), etc. Une donnée brute en tant que telle n'a pas vraiment de valeur, c'est en l'analysant et en la transformant qu'on peut l'utiliser pour prendre des décisions.

L'initiative Big Data s'appuie sur ces trois piliers pour répondre à des besoins, des enjeux, et des problématiques d'affaires qui seront considérés selon le cas d'usage.³

Ce qui est primordial à comprendre c'est qu'un projet Big Data ne concerne pas que les technologies. Il s'agit bien d'un tout qui englobe aussi les données, les pratiques, l'organisation et la gestion de projet, la gestion du changement, la gouvernance des données, etc.

¹Zoom sur les données : Les données Big Data sont souvent caractérisées par les 3 ou 5 V dont les plus importants sont la vitesse à laquelle les données sont produites, intégrées, transformées, et utilisées. La volumétrie ou le volume des données à stocker et à utiliser, et la variété des données. Une donnée provenant d'une base de données opérationnelle est structurée alors qu'une bande audio ou un enregistrement vidéo sont non structurés. Tous ces facteurs compliquent la tâche aux systèmes patrimoniaux des entreprises.

²Zoom sur le NoSQL : Les bases Not only SQL ont été conçues pour stocker et procéder à des requêtes sur des données qui ne peuvent pas être modélisées de façon tabulaire traditionnelle. Il existe plusieurs formats NoSQL et les plus populaires sont les modèles clé-valeur (key-value stores), les modèles en colonnes (column stores), les modèles orientés documents ou objets (document ou object databases), et les modèles orientés graphes (graph databases). Ce qu'il faut retenir c'est que pour un cas d'utilisation précis et un type de données précis, il existe probablement une technologie appropriée pour en venir à bout.

³Sans entrer dans les détails techniques, ces algorithmes nécessitent généralement une puissance de calcul que les technologies Big Data supportent. Ces algorithmes sont souvent vieux. Ce qui est nouveau, c'est le fait de pouvoir faire des calculs sur des pétaoctets de données. Les statisticiens étaient jadis limités, car aucun ordinateur sauf les supers ordinateurs n'était en mesure de traiter autant de données. L'intérêt d'utiliser autant de données est dans l'exactitude des résultats. Prendre l'ensemble des observations d'un phénomène amènera toujours des résultats plus précis que d'utiliser un échantillon de ces résultats et aujourd'hui, avec les technologies Big Data, il n'est plus nécessaire de s'en tenir à un échantillon.

POURQUOI SE LANCER DANS LE BIG DATA ?

Bien sûr, ce n'est pas parce qu'une entreprise se rend compte qu'elle produit beaucoup de données qu'elle doit soudainement les capter et les stocker. La tentation de démarrer un projet de mégadonnées est parfois forte : le terme à la mode est sur toutes les lèvres, l'idée semble novatrice et la haute direction veut bien paraître alors elle demande à l'équipe de TI de s'y mettre au plus vite. Les raisons sont nombreuses. En réalité, un projet informatique représente des coûts, et ces coûts doivent être rentabilisés, sinon ce n'est plus un investissement. Étonnamment, cette règle semble disparaître quand il s'agit de nouvelles pratiques aussi médiatisées que les mégadonnées.

Toutefois, une initiative Big Data peut s'avérer pertinente quand l'entreprise, d'un point de vue technique, opérationnel, ou stratégique :

- » n'arrive pas à répondre aux demandes d'analyse de données des différentes unités d'affaires, soit parce que l'accès est difficile ou parce que les traitements sont très longs ;
- » a de la difficulté ou n'arrive pas à acquérir, à stocker ou à traiter certaines données qu'elle produit ou qui viennent de l'extérieur de l'entreprise ;
- » a de mauvaises performances avec ses bases et entrepôts de données en place ;
- » subit un changement dans le paysage concurrentiel de son domaine d'affaires et n'arrive plus à suivre.

JE NE PEUX PAS FAIRE ÇA...

« Il paraît que les données machine (« logs ») sont intéressantes à analyser mais selon leur format, on ne peut pas les mettre dans nos bases SQL... »

C'EST QUOI CES DEMANDES...

« Les divisions voudraient profiter d'un apprentissage automatique, mais on n'a pas les outils pour le faire... »

ÇA FONCTIONNE MAL...

« Les TI aimeraient bien vous fournir les rapports plus rapidement, mais le traitement prend plusieurs jours à s'effectuer... »

COMMENT ARRIVENT-ILS À FAIRE ÇA ?

« Le concurrent fait plus de ventes que nous, mais on ne comprend pas comment il y arrive... »

En d'autres termes :

Ces situations se produisent parce que la nature des données a évolué avec le temps. Les systèmes d'informations patrimoniaux ont été implantés avec des objectifs précis, pour répondre à des besoins précis, et résistent souvent mal aux nouveaux défis.

Les impacts de telles restrictions sont souvent vus comme de petits désagréments, comme un mal nécessaire, mais en réalité, une entreprise qui est limitée par son système d'information rate des occasions qui peuvent changer considérablement les activités et les finances de l'organisation. Surtout lorsqu'il s'agit de données, l'avantage revient à celui qui l'exploite et dans un marché compétitif, les conséquences peuvent être néfastes pour l'entreprise qui ne s'est pas intéressée au sujet.

Bien sûr, les technologies Big Data ne sont pas la réponse miracle à tous les besoins informatiques, mais lorsqu'il s'agit de données, et encore une fois de la nature de ces données, les technologies phares sont généralement une bonne solution pour relever ces défis.

Start small, think big!



L'IMPACT SUR L'ENTREPRISE

Mais quel est le réel impact sur l'entreprise ? Dans le cas d'un projet Big Data, le rendement du capital investi peut se faire de différentes façons : une hausse des revenus de l'entreprise, une baisse des dépenses liées à l'exploitation, ou même une amélioration du fonctionnement de certaines parties de l'entreprise, ce qui est plus difficile à quantifier.

REVENUS

Améliorer les opérations de l'entreprise

Créer de nouveaux produits ou services

DÉPENSES

Réduire les frais informatiques

Renforcer la sécurité

LA HAUSSE DES REVENUS

Lorsque l'on parle de hausse de revenus, c'est qu'une décision d'affaires ayant une incidence financière positive a été prise, basée sur des données qui ont été analysées. C'est plutôt simple, non ?

Il est possible de mettre les cas Big Data liés à la hausse de revenus dans deux catégories :

1. L'amélioration des opérations de l'entreprise

L'amélioration des opérations peut être divisée en deux volets : l'aide à la décision humaine, dans le même style que l'intelligence d'affaires plus traditionnelle ; et la prise de décision automatisée, le genre qui interagit directement avec un client, sans intervention humaine, par exemple.

L'aide à la décision n'est pas nécessairement ce qu'il y a de plus novateur, mais ce qui l'est, c'est de pouvoir croiser différentes sources de données et traiter l'ensemble de ces données, qu'elles fassent un gigaoctet ou cent téraoctets. Il n'y a pratiquement plus de limites à l'analyse de données.

Voici quelques exemples :

- » Des données dites « froides » ou historiques qui sont normalement entreposées hors des systèmes décisionnels en raison de leur âge et de leur volume peuvent être utilisées en entier pour affiner les analyses des opérations et mieux comprendre l'évolution de l'entreprise à long terme.
- » Une meilleure gestion de l'inventaire est possible grâce au croisement des données de ventes historiques aux données de distribution, et même avec l'ajout de données météorologiques. On se doute que la vente de climatiseurs est en hausse avec l'augmentation de la température, mais est-ce vrai, et peut-on réagir de manière à trouver le juste équilibre entre la rupture de stock et l'excès en stock ?
- » Grâce à une analyse des différentes étapes d'un processus manufacturier, il est possible de mieux le comprendre et de l'améliorer et par le fait même, de réduire le gaspillage, d'augmenter la cadence, d'améliorer la synchronisation avec d'autres, etc. L'idée est toujours de faire mieux en s'adaptant à la réalité mesurée.

La prise de décision automatisée est une belle innovation qui permet de faciliter notamment les contacts avec les clients et malgré qu'il n'y ait pas d'intervention humaine, de lui fournir facilement et rapidement un service pertinent.

Les exemples sont abondants et variés, mais les plus connus sont :

- » Les assistants virtuels *chat bots* qui permettent de répondre à des questions de clients, de conclure des transactions, d'effectuer des opérations pour les utilisateurs, etc. Ces assistants réussissent à être performants d'une part grâce à ce qu'ils ont appris sur une grande quantité de données et d'autre part parce que lorsqu'un utilisateur pose une question, des calculs se produisent en arrière-plan en fonction du contexte de l'utilisateur.
- » La reconnaissance de données multimédias est aussi en vogue et utile, par exemple, en ce qui a trait à la reconnaissance et à la conduite assistée par ordinateur. Ces techniques nécessitent une grande quantité de données, en l'occurrence des photos et des vidéos, afin d'apprendre à reconnaître des visages, des routes, des panneaux de signalisation, etc.

2. La création de nouveaux produits ou services

En analysant des données internes et externes, l'entreprise est en mesure de créer ou d'affiner un produit ou un service qui pourra mieux répondre aux besoins de ses clients et qui aura de meilleures chances de succès. C'est une façon pour une entreprise d'être proactive et novatrice plutôt que d'attendre les nouvelles tendances. De plus, l'amortissement du développement du produit ou service est plus sûr.

Par exemple :

- » On voit apparaître de nouveaux modèles d'affaires dans le domaine de l'assurance automobile où la prime d'assurance est calculée en fonction de l'utilisation du véhicule par l'assuré. Cette technique n'est possible que par l'utilisation de technologies et de processus décisionnels qui sont centrés sur la collecte et l'utilisation de données massives.
- » Une entreprise peut utiliser les attributs clés de ses produits qui ont eu du succès pour en créer un nouveau. Cette dernière doit intégrer une multitude de sources de données (paniers d'achat, distribution, efforts marketing, etc.) pour bien comprendre les différents facteurs d'influence, enrichir ses analyses et s'assurer que ses résultats sont statistiquement valides.
- » L'acquisition et le traitement de données provenant des réseaux sociaux peuvent permettre à une entreprise de comprendre ce que son public cible apprécie de ses produits et de ceux de ses concurrents. C'est une mine d'or de données qui peuvent servir à la création de nouveaux produits ou services.

Une bonne partie des grandes entreprises présentent sur le Web et des entreprises en démarrage basent non seulement leurs activités sur le traitement de données, mais aussi leurs offres de services sur ces mêmes analyses de données.

LA BAISSÉ DES DÉPENSES

Tout comme pour les cas de hausse de revenus, la baisse des dépenses s'effectue généralement au moyen de deux moteurs :

1. La réduction des frais informatiques

Il s'agit de la réduction des dépenses liées à l'exploitation, spécialement en ce qui concerne les frais informatiques. En d'autres mots, payer moins cher pour plus de performances. Sans rentrer dans les détails techniques, la raison est que de nombreuses technologies Big Data sont ouvertes (*open source*) et qu'elles ne nécessitent pas de matériel informatique spécialisé. Cela peut se traduire entre autres par le remplacement d'un outil payant, sous licence, par un outil de type logiciel libre.

2. Le renforcement de la sécurité

Le renforcement de la sécurité couvre plusieurs aspects, mais dans l'ensemble, l'idée est simple : comment l'entreprise peut-elle éviter les pertes en analysant de façon proactive les données de ses systèmes d'information ?

Par exemple :

- » Analyser l'ensemble du trafic sur son site transactionnel avec les habitudes des clients afin de déceler les risques de fraude, comme l'usurpation d'identité.
- » Analyser les données machines (*logs*) afin de détecter une potentielle intrusion dans le système informatique de l'entreprise.
- » Extraire les signaux faibles des données d'activité de l'entreprise afin de déceler des cas de fraude ou des activités anormales pouvant avoir un impact négatif sur l'entreprise.

Tous les exemples énoncés, que ce soit pour la hausse des revenus ou la baisse des coûts, sont des situations où l'analyse de données – et en toute honnêteté, pas nécessairement des mégadonnées – peut mener à des résultats positifs pour l'entreprise. Il peut s'agir d'un simple événement qui, lorsqu'il se produit, entraîne une action lorsque la donnée est créée.

Certains cas sont moins évidents que d'autres en ce qui a trait aux répercussions, mais ce qu'il faut garder en tête, c'est que s'il y a une prise de décision basée sur des données, il y aura probablement une conséquence positive qui devrait être déterminée et pérennisée.

SANS BESOINS D'AFFAIRES, OUBLIEZ LE BIG DATA !

Jusqu'ici, nous avons présenté les raisons de se lancer dans une initiative Big Data. Cela peut sembler simple, mais à vrai dire, ce genre d'initiative n'est jamais gagné d'avance. Tout gestionnaire est d'accord pour se donner les moyens d'augmenter les revenus de son entreprise et de diminuer ses coûts. Sans aborder les technologies, comment peut-on faire ? Le seul moyen, c'est de définir les besoins d'affaires de son entreprise, avant toute chose. En d'autres termes, quels sont les objectifs précis que l'entreprise souhaite atteindre par l'acquisition, le stockage et le traitement de données massives ?

UN BESOIN D'AFFAIRES ?

Dans notre contexte, un besoin d'affaires, une exigence d'affaires ou un enjeu d'affaires sont des synonymes de cas d'usage ou de cas d'utilisation. Autrement dit, un cas d'utilisation Big Data est une suite d'événements et de traitements de données qui ont au moins un objectif précis et qui ont des répercussions positives sur l'entreprise c'est-à-dire une hausse des revenus ou une baisse des coûts.

Selon IBM, un cas d'usage permet à l'entreprise de résoudre un défi d'affaires précis, en utilisant des modèles ou des exemples de solutions technologiques qui ont fait leurs preuves. Il est donc important de bien les définir et de bien les documenter pour déceler tout intérêt individuel ou collectif qui ne serait pas pertinent pour la réussite du projet. Autrement dit, il faut s'assurer que notre initiative est réellement rentable.

Les entreprises ont généralement certains défis à relever, des problèmes à régler et des processus non optimaux à améliorer. Ce sont souvent ces défis qui forment les besoins d'affaires sur lesquels on peut se baser. Et la question à se poser est la suivante : comment puis-je répondre à ces besoins, à ces cas d'utilisation avec des technologies appropriées ? Le concept n'est pas nouveau, mais il est toujours aussi pertinent dans un contexte d'initiative Big Data :

« Dis-moi ce que tu veux faire et je te dirai ce qu'il te faut pour y arriver ».

Il est primordial de réfléchir avant d'agir et de ne pas commencer par la technologie. En effet, depuis que les entreprises se sont informatisées, le défi a toujours été de relier les technologies aux besoins et aux processus d'affaires de l'entreprise. Pour diverses raisons, dont la popularité du domaine, lorsque l'on parle de technologies Big Data, on observe souvent les gestionnaires s'y prendre à l'envers : « Qu'est-ce que je vais faire avec mes technologies Big Data fraîchement implantées ? »

Il a été maintes fois prouvé que ce genre de démarche ne donne pas les résultats escomptés parce qu'il y a souvent peu de besoins définis (donc on n'est pas certain d'avoir un taux de rentabilité suffisant), et même si l'on détermine des besoins après l'implantation des technologies, ces dernières n'ont peut-être même pas les fonctionnalités nécessaires pour répondre aux besoins ! En bref, le seul et unique point de départ d'une initiative Big Data est le besoin d'affaires, pas la technologie ! Qu'il soit déjà défini ou non, un besoin se développe et s'affine assez facilement avec l'accompagnement, l'expertise et les méthodologies appropriés.



Voici un exemple de chaîne de valeur d'une entreprise manufacturière :



LES BESOINS, ÇA PART D'OÙ ?

Toute entreprise qui s'intéresse aux mégadonnées doit commencer quelque part et la meilleure manière de commencer est de s'intéresser aux activités principales de l'entreprise : les secteurs d'activités. Les secteurs d'activités, comme le marketing, la production ou les finances, sont eux-mêmes composés de processus d'affaires, et tous ensemble, ils forment la chaîne de valeur de l'entreprise.

Lorsque l'on s'intéresse au phénomène Big Data, on souhaite d'abord et avant tout avoir une incidence positive sur les activités principales de l'entreprise par l'analyse de données. Comme nous l'avons mentionné, on souhaitera aussi avoir un effet positif sur les technologies de l'information même si ces dernières sont généralement considérées comme étant une activité de soutien à l'entreprise. Il ne faut pas se limiter ; s'il y a un moyen d'exploiter des données et d'en tirer des bénéfices, l'origine n'est plus importante. Dans les activités de soutien, nous retrouverons généralement des initiatives d'analyse de données qui permettent de réduire les coûts de l'entreprise et dans les activités principales, l'analyse de données Big Data se traduira la plupart du temps par une potentielle hausse des revenus ; quoique ce soit non exclusif.

L'objectif est d'observer et d'analyser les processus de l'entreprise qui se fondent sur des données (ou non) et de voir comment il serait possible de bonifier cette prise de décision, voire de l'automatiser et de la rendre plus performante. En d'autres termes,

« Quelles décisions sont prises dans mon entreprise et comment puis-je améliorer cette prise de décision par des chiffres concrets ? »

En fin de compte, le besoin n'a pas à être clair et net, tout ce qu'il faut c'est une idée, un petit ou grand problème qu'on pourrait améliorer, et à partir de là tout s'enchaîne.

Voici quelques exemples de processus d'affaires qui peuvent être améliorés grâce à l'analyse de données :

- » Une entreprise a l'habitude d'effectuer l'entretien de ses camions aux deux cent mille kilomètres. Une analyse des carnets de bord et des réparations effectuées permettrait-elle de prévenir des pannes et d'améliorer l'entretien de véhicules ?
- » Une entreprise sollicite ses clients qui font au-delà de cent transactions par année en leur faisant part d'offres particulières pour essayer de mousser ses ventes. Est-ce qu'une analyse de l'ensemble des transactions de tous ses clients confondus ne lui permettrait pas de cibler de potentiels acheteurs qu'elle n'aurait pas repérés par leurs volumes d'achats ?
- » La société mère d'un grand magasin aimerait mieux connaître les caractéristiques socio-économiques de ses clients et comprendre les tendances d'achat, autant dans ses succursales qu'en ligne, afin de mieux adapter son effort marketing.

DÉTERMINER SES BESOINS D'AFFAIRES

Nous le savons déjà, les cas d'usage ou les besoins d'affaires sont à la base de toute initiative Big Data. Notre expérience nous a démontré que ce n'est pas parce qu'une entreprise implante des technologies Big Data qu'elle exploitera adéquatement les données qu'elle stocke. C'est le grand classique, une technologie prometteuse pour l'entreprise n'est finalement pas toujours utilisée. N'est-ce pas du déjà vu ? Dans tout projet informatique, il est nécessaire de voir un certain taux de rentabilité à moyen ou à long terme, sinon c'est une perte nette pour l'entreprise. Les projets de mégadonnées ne devraient pas être l'exception à la règle.

Donc, qu'est-ce qui est nécessaire, avant même les technologies, pour qu'une entreprise puisse se lancer avec confiance dans une initiative Big Data, réussir, et ne jamais regarder en arrière ?

Ce qui est nécessaire, c'est de définir ses propres cas d'usage et les besoins propres à son entreprise (ou son secteur d'activités) et d'évaluer s'il y a un potentiel de rentabilité. Une approche complémentaire est de simplement s'approprier les cas d'usage de référence qui ont déjà fait leurs preuves. Par contre, il s'agit d'une approche moins attrayante parce que les risques de résistance au changement sont plus grands lorsqu'une équipe doit intégrer quelque chose qu'elle n'a pas nécessairement choisi ni développé. Nous verrons ces cas de référence au prochain chapitre.

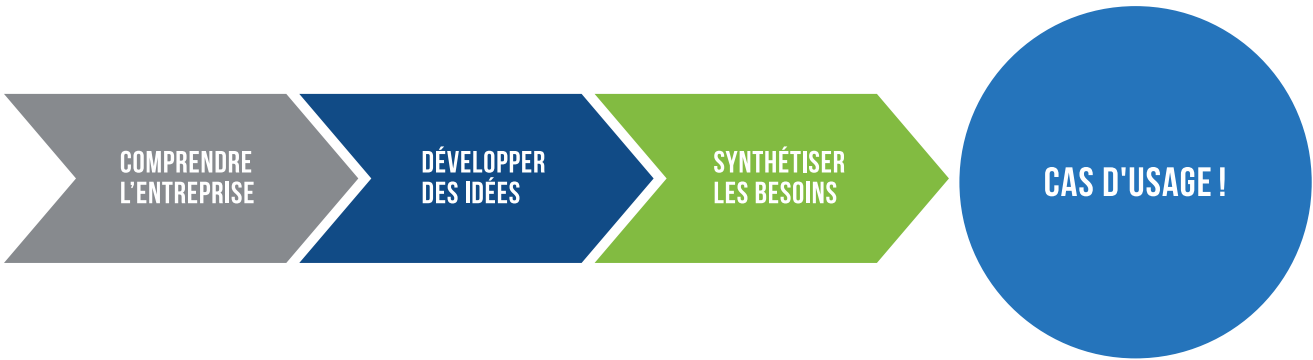
UNE MÉTHODOLOGIE DE DÉVELOPPEMENT

Il n'est pas facile de déterminer les étapes à suivre et le meilleur moyen d'y arriver est d'avoir recours à une approche structurée, accessible et non contraignante, qui laisse place à la créativité, au développement d'idées et à l'innovation.

La plupart des étapes de développement d'un projet Big Data n'ont rien de bien nouveau ; ce sont des processus assez traditionnels de planification et d'organisation. Ce qui est nouveau, mais pas toujours respecté, c'est entre autres de prendre réellement le temps de développer des cas d'usage pertinents.

NOVIPRO propose une méthodologie simple et éprouvée pour développer des cas d'usage en faisant appel à un grand nombre d'intervenants par l'entremise d'une méthode structurée. Cette méthodologie est accessible à tous, mais il est recommandé de choisir un responsable de l'initiative qui en assurera la direction. Ce dernier doit avoir une expertise en matière de Big Data et savoir diriger des entretiens et des débats d'idées en groupe.

Les trois grandes phases de la méthodologie sont : comprendre l'entreprise, développer des idées et synthétiser les besoins.



COMPRENDRE L'ENTREPRISE

Il s'agit d'abord d'étudier, dans une certaine mesure, l'entreprise et le contexte d'affaires dans lequel elle évolue. Tout comme si l'entreprise souhaitait préciser sa stratégie d'affaires, il est pertinent de bien saisir les différents contextes internes et externes de l'entreprise.

ÉTAPES	DESCRIPTIONS
1	Comprendre les processus d'affaires (et de soutien) Étudier les processus qui procurent une valeur à l'entreprise ainsi que les données utilisées (ou pas) par ces processus. À court terme, ce sont ces processus facilement identifiables qui peuvent bénéficier d'une initiative Big Data.
2	Faire un inventaire des données produites Il est important de connaître l'ensemble des données produites par l'entreprise, les contextes dans lesquels elles sont produites, qui les utilise, quelles sont les règles de gestion, etc. Le but est d'avoir une bonne vue d'ensemble pour pouvoir évaluer les forces et les faiblesses de l'entreprise en ce qui concerne l'accès aux données.
3	Déterminer les défis en matière de fonctionnement interne et externe L'idée est de voir, sommairement, où sont les problèmes de l'entreprise. Il ne faut pas se limiter aux processus qui utilisent des données. Il faut plutôt regarder l'ensemble des blocages et des défis, car ces derniers peuvent bénéficier d'une analyse de données.
4	Rassembler les idées préliminaires d'amélioration Au moyen d'entrevues avec différents intervenants, il est possible de cibler rapidement des idées d'analyse de données. On ne vise pas ici la qualité, mais bien la quantité d'idées d'amélioration ; encore une fois, qui ne sont pas nécessairement axées sur l'utilisation de données.

Que ce soit pour les processus, les données, les défis, ou les idées, n'importe qui peut participer et donner son avis sur ces sujets ; du président de l'entreprise jusqu'à l'employé d'entrepôt, tous ont le potentiel d'avoir des solutions à des problèmes. C'est pourquoi des entretiens peuvent et devraient être réalisés au besoin avec n'importe quelle partie prenante.

En somme, il est important de ne pas se limiter dans son étude de l'entreprise et de ne pas avoir d'idées préconçues. De plus, en fonction de la taille de l'entreprise comme dans une PME, il peut être réaliste de faire appel à l'ensemble des secteurs d'activités, alors que dans une grande entreprise, la seule façon viable sera de commencer par un secteur d'activités bien précis.

DÉVELOPPER DES IDÉES

Il s'agit de mettre les différents intervenants de l'entreprise, les parties prenantes, sur un pied d'égalité pour ensuite les amener à réfléchir à des problèmes, des défis, des solutions ou des besoins, qu'ils pourraient avoir dans leur travail de tous les jours.

L'expérience a démontré que les gens n'ont pas nécessairement un manque d'inspiration quand il s'agit de développer des cas d'usage Big Data, mais lorsque ces derniers ne connaissent pas l'étendue des possibilités qu'offre le Big Data, il est plus difficile d'exprimer des idées sur le sujet. Les différents secteurs d'activités ont en effet tendance à se limiter dans l'expression de leurs besoins en exprimant des idées qui cadrent avec ce qu'elles considèrent comme réalisable, sans connaître l'étendue des possibilités technologiques !

Cette séance est donc une excellente occasion de présenter ce qui se fait dans l'industrie en ce qui concerne les cas d'usage Big Data, toujours dans l'idée d'ouvrir l'esprit aux gens et de leur permettre de mieux réfléchir à ce qu'il serait pertinent d'appliquer à leur propre entreprise.

ÉTAPES	DESCRIPTIONS
5	Former les parties prenantes (et même plus) à l'état de l'art et aux concepts Big Data Que ce soit les TI, les secteurs d'activités, les activités de soutien, la direction ou les analystes, tous n'ont pas les mêmes connaissances du domaine Big Data. Il est donc primordial de former sommairement l'ensemble des intervenants afin de favoriser le développement et l'innovation en matière de besoins d'affaires. Comme nous l'avons déjà mentionné, les secteurs d'activités ont tendance à se limiter dans leurs besoins. Leur exposer la réalité du Big Data, dont les possibilités qu'offrent les trois piliers, permet d'éveiller l'imagination. Ces périodes de formation sont aussi une excellente occasion de mettre à jour les différentes constatations et découvertes faites à la première phase. Il s'agit donc de se donner tous les outils possibles pour préparer adéquatement la prochaine étape.
6	Animer des ateliers de remue-méninges pour formuler des besoins d'affaires Concrètement, il s'agit d'échanger, en groupes variés, des idées relatives à l'amélioration (ou même à la création) de processus d'affaires par une analyse de données, l'aide à la prise de décision, l'automatisation, etc. Nul n'est mieux placé pour réinventer son métier que l'expert de son domaine qui devra lui-même s'approprier les nouvelles façons de faire. Tous peuvent être invités à ce genre d'activité. Certaines personnes sont bien sûr susceptibles d'avoir des idées plus pertinentes que d'autres, mais on ne sait jamais qui dans l'entreprise a pensé à une solution, à un processus qui pourrait être amélioré, ou même qui utilise des données qui pourraient l'aider dans son travail. L'idée est de ne pas se limiter. Une séance de remue-méninges bien organisée et dirigée, où les participants ont une vraie expertise de leurs domaines d'affaires et où ils ont été sensibilisés aux concepts Big Data (étape 5), a tout le potentiel de produire des cas d'usage et besoins d'affaires qui ont une réelle valeur ajoutée.

Nous proposons ici une activité de remue-méninges, mais plusieurs autres techniques de développement d'idées peuvent être mises en application, comme la conception créative (*design thinking*), le scénarimage (*storyboarding*), la pensée inversée (*reverse thinking*), etc. Il n'y a pas de bonnes ni de mauvaises façons de développer des idées, il faut simplement réunir les intervenants et prendre le temps de bien le faire.

Il y a plusieurs avantages à faire cet exercice une ou plusieurs fois avec des groupes différents et des groupes mixtes. Premièrement, l'entreprise ne se retrouvera pas à faire comme toutes les autres parce que c'est ce que les autres entreprises font ; ce seront des idées et des besoins qui seront calqués sur la réalité de l'entreprise. Deuxièmement, lorsque les gens définissent eux-mêmes leurs besoins et se les approprient, lors de l'implantation technologique, les risques d'échec dû à la résistance au changement sont réduits parce que l'on n'impose pas de nouvelles façons de faire aux gens ; les idées viennent des utilisateurs.

La deuxième phase est potentiellement longue, c'est-à-dire qu'elle dure souvent plusieurs semaines puisqu'il faut réunir les intervenants, mais cela demeure essentiel. L'entreprise qui va trop vite et qui adopte une technologie avant d'avoir fait le tour de la question et des besoins risque de ne pas voir cette technologie adoptée par ses utilisateurs.

SYNTHÉTISER LES BESOINS

Il s'agit de consolider l'ensemble des informations recueillies et de les traduire en besoins d'affaires ou en cas d'usage concrets et réalisables.

ÉTAPES	DESCRIPTIONS
7	<p>Rassembler les différents besoins et les idées exprimées et les transformer en cas d'usage concrets</p> <p>Concrètement, qu'est-ce qui a été exprimé, qu'est-ce que cela sous-entend et comment peut-on y répondre avec l'aide des trois piliers Big Data ? Est-ce que les données sous-jacentes existent ? Comment répond-on à ce besoin ?</p>
8	<p>Évaluer le potentiel de rentabilité de ces cas d'usage</p> <p>Un cas d'usage doit avoir une incidence financière positive. Est-ce possible de l'évaluer avant de se lancer dans l'aventure ? Si oui, est-ce que son apport financier est substantiel et assez intéressant pour motiver (en partie) un investissement technologique ?</p>
9	<p>Sélectionner quelques cas d'usage prometteurs</p> <p>Généralement, quelques cas d'usage variés peuvent motiver un investissement dans une technologie Big Data. Ces derniers ont assez de valeur pour les secteurs d'activités en ce qui concerne l'aide à la prise de décision, l'innovation, l'amélioration des opérations, etc.</p>
10	<p>Facultatif : communiquer les cas d'usage</p> <p>Une fois la sélection faite, il peut être pertinent, quoique redondant, de partager les cas ciblés avec les gens impliqués dans le processus de développement pour préciser le besoin et s'assurer qu'il a été bien compris.</p>

C'est à partir de ce moment que l'entreprise peut évaluer si oui ou non il y a un avantage financier à se lancer dans une initiative Big Data puisque des besoins clairs et réalisables ont été déterminés et que s'ils sont comblés, ils ont le potentiel d'améliorer les performances de l'entreprise. Les solutions à ces besoins ne sont pas automatiquement des technologies de type Big Data ou des données supplémentaires à capturer, mais dans tous les cas, l'exercice met l'entreprise sur la bonne voie de l'amélioration !

La troisième phase peut se réaliser dans un temps raisonnable, en fonction de l'ampleur de la tâche de conciliation.

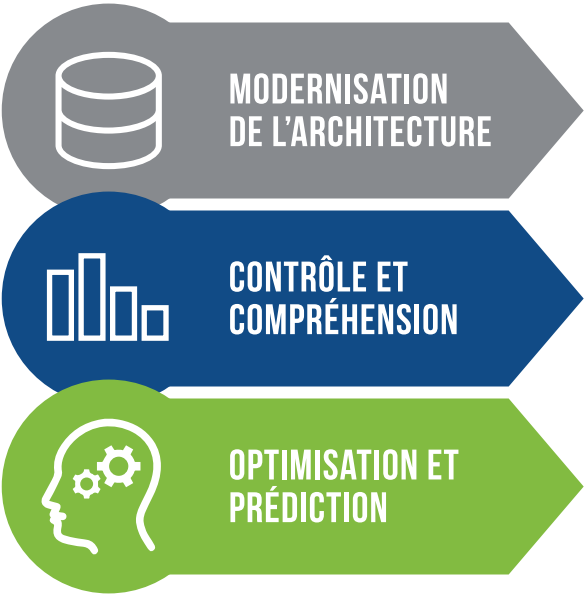
LES CAS D'USAGE DE RÉFÉRENCE

Jusqu'ici, nous avons vu que les technologies Big Data doivent essentiellement répondre à un besoin d'affaires bien défini.

On a exploré une méthodologie assez simple et structurée pour déterminer ses besoins d'affaires. Mais qu'en est-il de ce qui se fait généralement sur le marché ? Y a-t-il des cas qui sont tellement fréquents qu'une entreprise ne peut passer à côté, même si de prime à bord elle ne les définit pas comme étant des besoins ? La réponse est oui.

Ces cas ont déjà démontré leur valeur sur le marché. Les éditeurs de technologies Big Data ont été rapides pour les reconnaître, d'une part pour leurs besoins de communication et de marketing, mais surtout parce que les entreprises qui s'y sont attelées ont vu un réel rendement sur leur investissement.

Il est difficile de mettre l'ensemble des projets Big Data dans des catégories très précises, mais nous en avons décelés trois, qui peuvent se chevaucher, mais qui en fin de compte, permettent de bien comprendre l'étendue de ces cas.



- OPTIMISATION DE BASE DE DONNÉES (SGBD)
- DÉCHARGEMENT ETL
- ARCHIVAGE ACTIF

- RAPPORTS D'ACTIVITÉ OPTIMISÉS (BI+)
- DÉCOUVERTE, EXPLORATION ET ENRICHISSEMENT
- ANALYSE DES PROCESSUS / PARCOURS

- VUE UNIQUE (360° VIEW OF CUSTOMER)
- RECOMMANDATION, SEGMENTATION, ETC.
- MAINTENANCE PRÉDITIVE, DÉTECTION DE LA FRAUDE, DÉTENTION DE SIGNAUX FAIBLES

Les tableaux suivants regroupent les cas d'usage les plus fréquents dans les entreprises qui se sont lancées dans une initiative Big Data. Bien sûr, ce n'est pas parce que le compétiteur a mis en production un cas d'usage qu'il est adapté à l'entreprise, mais c'est un bon indice de ce qui devrait intéresser certains secteurs d'activités.

MODERNISATION DE L'ARCHITECTURE

CAS D'USAGE	BESOINS D'AFFAIRES	↑ \$	↓ \$
Entrepôt de données optimisé	Accéder à l'ensemble des données historiques, en provenance de plusieurs sources, pour avoir des analyses plus complètes.	×	
Archivage actif	Accéder à l'ensemble des données historiques, de plusieurs sources, pour bonifier les analyses. Les entrepôts de données contiennent rarement plus que quelques années de données, ce qui pénalise les analyses sur l'historique.	×	
Alimentation embarquée	Remplacer ou améliorer les outils d'alimentation et d'extraction de données (ETL) grâce à de meilleures performances et à des capacités de gestion de la volumétrie, de la variété et de la vitesse des données.	×	
Collecte de données disparates	Capacité d'acquérir des données internes ou externes à l'entreprise, peu importe leur forme : données Internet des objets (IdO), données machines (logs), données publiques ou données audio, vidéo ou texte, etc.	×	
Réduction des dépenses d'exploitation et du coût total de possession	Réduction des frais comme ceux liés aux licences sur une vaste gamme de technologies : bases de données (SGBD), outils de collecte (ETL), outils de traitement statistique des données, outils de visualisation de données, connecteurs et autres. Il n'est pas rare de voir des éditeurs historiques de logiciels statistiques se faire remplacer par des technologies plus performantes et, ouvertes !		×
Déchargement de l'ordinateur central	Diminuer la quantité de données historiques stockées dans l'ordinateur central (main-frame) et faciliter l'accès à ces données pour les analystes.		×
Indexation et recherche	Permettre la recherche sur l'ensemble des mots d'une multitude de formats de fichiers : CSV, PDF, Word, etc.	×	
Lac de données (données-service)	Briser les silos de données de l'entreprise en stockant à un même endroit les données en provenance de multiples silos de l'entreprise. Ce qui permet une analyse de données enrichie.	×	

CONTRÔLE ET COMPRÉHENSION

CAS D'USAGE	BESOINS D'AFFAIRES	↑ \$	↓ \$
Rapport d'activités	Améliorer l'intelligence d'affaires grâce à des capacités de traitement plus rapides et efficaces, sur un volume de données historiques quasi illimité.	×	
Découverte et exploration	Analyser des données, brutes ou non, sans jugement, pour découvrir des tendances et profiter de ces données. L'exercice est difficile avec des entrepôts ou des magasins de données traditionnels qui sont mis en forme selon des règles précises puisqu'il laisse peu de place à l'exploration.	×	
Analyse des opérations	Avoir un nouveau regard sur les opérations pour une optimisation des résultats et une réduction des coûts.	×	
Enrichissement de données	Bonifier les analyses avec des jointures de données.	×	×
Analyse de données disparates	Permettre de profiter des données qui ne sont pas toujours analysées comme : les données machine (logs), les appels téléphoniques (audio), les fichiers texte, les données de capteurs (IdO), etc.	×	
Cartographie sociale	Permettre une analyse des relations entre les données. Les analyses de type graphes souvent associées aux bases de données graphes, permettent de comprendre les interactions et les liens entre des agents, comme des individus.	×	
Suivi de paiement	Accélérer l'analyse des chemins de vente pour anticiper et potentiellement réduire les cas de fraude qui se répètent ou non.	×	
Analyse de panier de consommation	Être en mesure d'analyser l'ensemble des paniers de consommation pour déterminer les liens entre les produits achetés et ainsi favoriser la vente croisée.	×	
Analyse de parcours	Pouvoir suivre les parcours des clients sur l'ensemble des médiums de l'entreprise : site Web, succursales, centre d'appel, etc. Cela permet de comprendre les processus et de les améliorer.	×	



OPTIMISATION ET PRÉDICTION

CAS D'USAGE	BESOINS D'AFFAIRES	↑ \$	↓ \$
Vue unique (vue 360°)	Améliorer le suivi et l'engagement en rassemblant en un seul endroit l'ensemble des informations disponibles à propos d'un client ou un fournisseur (dont les cartes de pointage). Cela permet de mieux les connaître, de favoriser de bonnes relations et d'améliorer la prise de décision.	×	×
Service client	Améliorer le service à la clientèle grâce à l'automatisation : assistants virtuels (chat bots), intelligence artificielle, données de capteurs (IdO), etc.	×	×
Analyse de sentiments	Connaître la réputation de l'entreprise grâce à une capture et à une analyse des données provenant de l'interne (service à la clientèle) et de l'externe (réseaux sociaux), typiquement grâce à une exploration de texte (textmining).	×	
Rétention client	Être en mesure de déceler qu'un client est susceptible de quitter l'entreprise pour un concurrent grâce à l'analyse de l'ensemble de ses comportements.	×	
Segmentation	Pouvoir mieux comprendre et cibler les familles de clients de l'entreprise afin d'ajuster le discours marketing et favoriser les ventes, la fidélité et la satisfaction.	×	
Moteurs de recommandation	Être en mesure de proposer à des clients des produits qu'ils sont susceptibles d'acheter en raison de leur historique de consommation.	×	
Conception de produits ou services	Réussir à cibler un nouveau produit ou service qui serait proposé aux clients grâce à l'analyse de données. En d'autres termes, peut-on ajouter de la valeur à nos produits et services par l'analyse de données ?	×	
Détection de signaux faibles / fraude	Utiliser des algorithmes spécialisés pour détecter des actions ou des transactions qui sortent de l'ordinaire sans nécessairement être extravagantes, ce qui ultimement vise à réduire la fraude en détectant les nouvelles façons de faire des fraudeurs.	×	×
Maintenance prédictive	Prédire les bris ou les accidents avant que ces derniers se produisent ; sur des machines, des véhicules, des appareils, des bâtiments, etc.	×	×

Il est presque impossible de mettre tous ces cas d'usage dans des boîtes et il faut parfois forcer un peu pour les faire entrer. Ce n'est donc pas surprenant, quand on regarde attentivement, de voir certaines répétitions et certains cas qui pourraient très bien être dans une autre catégorie. Malgré le raccourci, l'objectif est de démontrer que la majorité des cas sur le marché ont ultimement des objectifs assez similaires.

De plus, certains de ces cas s'appliquent à des secteurs d'activités précis, mais d'autres sont plus vastes et applicables à n'importe quelle entreprise qui produit et gère des données et qui souhaite en tirer de la valeur, en fonction du niveau de maturité de cette dernière. Avant de courir, il faut bien sûr apprendre à marcher. Bien qu'il y ait des cas extrêmement intéressants, ils ne sont pas toujours à la portée de toutes les entreprises, et de toutes les bourses. En effet, certains

cas plus complexes nécessitent une certaine expérience de l'entreprise pour ce type de projets. D'une part à cause de la complexité de réalisation, mais aussi pour la complexité technique, le matériel informatique, l'acquisition de données à travers les silos, etc. La faisabilité est donc à évaluer dans le contexte de son entreprise.

L'exploration des données, ce cas d'usage qui n'en est pas vraiment un... Lorsqu'une initiative Big Data devient mature et que l'investissement dans les technologies et les talents est rentable pour l'entreprise, certaines vont profiter de cet accès à des données variées pour les explorer, tout simplement. C'est-à-dire que sans savoir d'avance si certaines données ou traitements de données sont pertinents, on permet aux analystes d'explorer ce qui a été mis à leur disposition, sans aprioris. L'objectif de cette exploration est de tenter de trouver des tendances, des phénomènes, des corrélations, etc. Peut-être que dans ces données se trouve la prochaine pépite qui amènera une aide à la décision ou un avantage concurrentiel auquel on n'avait pas pensé ?

CONCLUSION

Ce livre blanc avait pour objectif de présenter au lecteur la manière pertinente de se lancer dans une initiative Big Data. Que ce soit par une hausse du chiffre d'affaires ou une baisse des coûts d'exploitation, il est primordial de rentabiliser un investissement Big Data. Cette rentabilisation se fait en répondant à au moins un besoin d'affaires que l'entreprise sera en mesure de combler grâce à ses technologies Big Data.

Il y a bien évidemment une foule d'autres aspects qu'il est primordial de prendre en compte dans le cadre d'un projet Big Data. Par exemple : Comment réaliser une preuve de concept Big Data ? Quels talents sont nécessaires ? Est-ce qu'un scientifique des données est nécessaire ? Comment organiser et gérer une équipe Big Data, l'agilité, le concept de l'exploration de données, la gouvernance des données, la communication, la résistance au changement, les couches technologiques pertinentes, etc. ?

Toutes ces questions viennent après et elles doivent être abordées lorsque l'entreprise aura défini ses besoins et ses cas d'usage. Malgré le fait que les analystes d'affaires répètent depuis des décennies que les technologies doivent être au service de besoins d'affaires clairs et nets, il est primordial que la tendance se maintienne, même si l'on dit autant de bien des technologies Big Data !

Finalement, se lancer dans l'aventure, ou du moins se lancer dans l'évaluation de cette aventure n'est pas quelque chose de compliqué. Il s'agit d'un processus de développement comme un autre, mais un accompagnement par des experts du domaine, à toutes les phases d'un projet, doit être envisagé pour prévoir le temps et l'investissement, déterminer des bénéfices potentiels plus rapidement, et optimiser le taux de rentabilité.



FRANÇOIS VIENNEAU BINETTE

CONCEPTEUR DE SOLUTIONS BIG DATA

Fort de plusieurs années d'expérience sur des projets Big Data en Europe, François se spécialise dans l'accompagnement technique et fonctionnel sur la grande chaîne de valeur Big Data ; de la vision stratégique d'une initiative, en passant par l'installation et la gestion de projet d'implantation Hadoop. S'adressant autant aux gestionnaires de lignes d'affaires qu'aux responsables TI, François vise à consolider et rentabiliser les efforts Big Data afin de favoriser le retour sur investissement tant attendu par les technologies Open Source. Il est également certifié administrateur et développeur sur l'écosystème Apache Hadoop.



NOVIPRO

www.NOVIPRO.com | www.e-spacesystem.com | france.e-shs.com

1 866 726-5353 | 1 844 377-7978

DELLEMC
PARTNER
PLATINUM