



Informatica™

Le grand manuel des Big Data

Guide pratique pour vous aider
à démarrer votre premier projet Big Data

Table des matières

Passage aux Big Data	03	3^e partie : Votre chaîne logistique Big Data lean	29
1^{re} partie : Préparation	04	Votre équipe	30
Ce que vous devez savoir	05	– Cinq leçons fondamentales pour renforcer l'esprit d'équipe	31
– Pourquoi la plupart des entreprises implémentent des projets Big Data	06	– Implémenter la gouvernance de données	34
– Pourquoi certains projets Big Data n'aboutissent pas	07	– Les compétences dont vous avez besoin et celles dont vous disposez déjà	36
– Comment assurer le succès de votre projet Big Data	09	– Comprendre les outils liés aux Big Data	38
Choisir un projet adapté	11	Vos processus	39
– Caractéristiques d'un projet adapté	12	– Les huit étapes vers les Big Data	40
– Réfléchir à son impact	14	Votre architecture	42
– Quelques exemples de projets Big Data tactiques	16	– Première étape : votre sandbox	43
2^e partie : Votre stratégie	17	– L'architecture Big Data idéale	44
Définir vos objectifs	18	Votre plan de projet	45
– Objectifs métiers	19	– Démarrer	48
– Objectifs informatiques	21	Étapes suivantes	49
Définir vos besoins en données	23	À propos d'Informatica®	50
– De quelles données avez-vous besoin ?	24		
– Cinq considérations clés en matière de données	27		

Conseil : cliquez sur un titre pour atteindre directement la section.



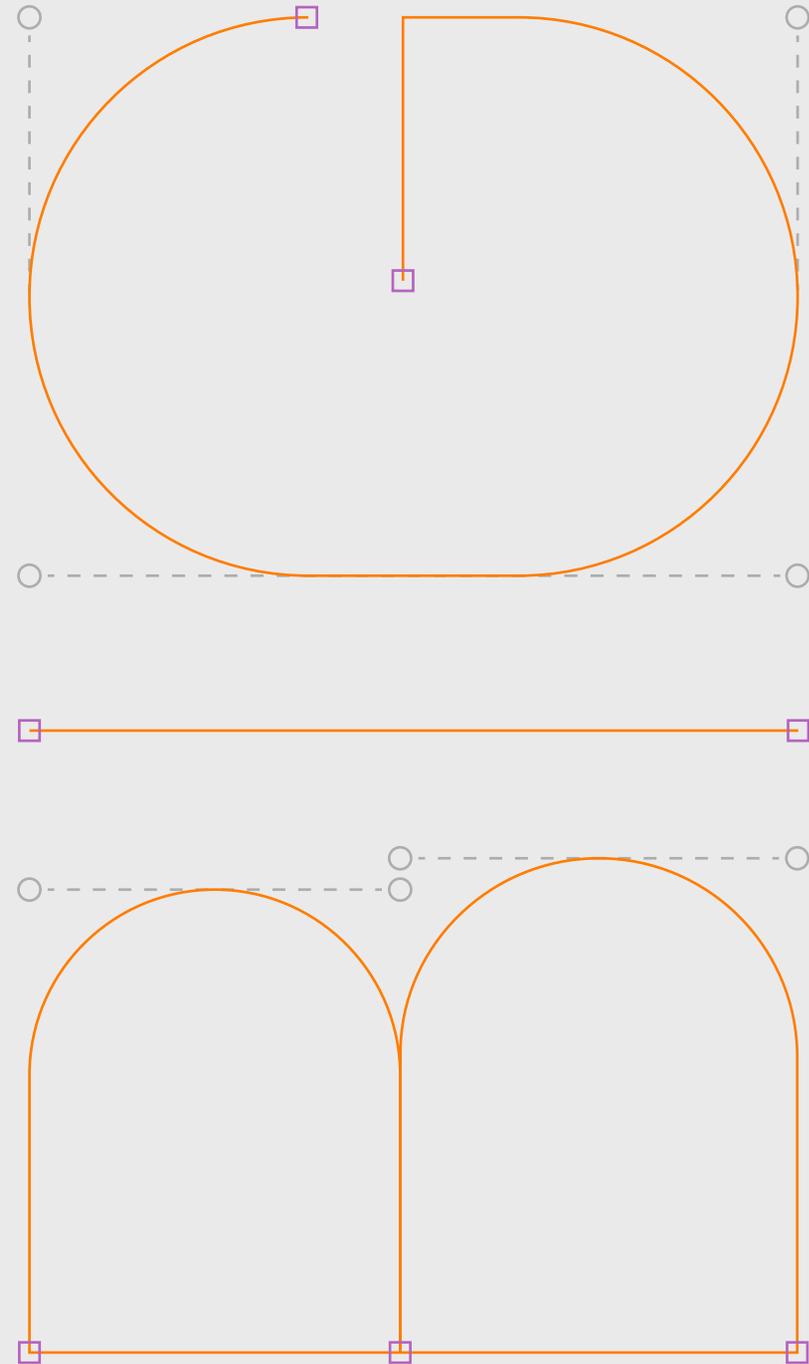
Passage aux Big Data

Très peu de technologies font autant parler d'elles que les Big Data.

Mais c'est probablement parce qu'elles sont peu nombreuses à offrir aux entreprises un potentiel de transformation si développé. Depuis que les logiciels ont commencé à englober des processus métiers entiers au début du siècle, une chose est claire : les données métamorphosent notre façon de travailler.

Que vous vous lanciez dans une initiative tactique localisée ou que vous planifiez un projet plus fondamental à l'échelle de l'entreprise, ce manuel vous guidera tout au long du parcours.

Entrons dans le vif du sujet.



1^{re} partie

Préparation

Nous avons divisé ce manuel en trois parties. Dans la première, nous allons essayer d'affiner votre vision pour vous aider à choisir un projet adapté.

Ce que vous devez savoir

Avant de vous pencher sur les aspects spécifiques de votre propre projet, découvrez les informations que la plupart des utilisateurs de Big Data auraient aimé connaître avant de démarrer leurs projets.



Pourquoi certains projets Big Data n'aboutissent pas

Selon une enquête¹, 55 % des projets Big Data n'aboutissent pas, et beaucoup d'autres n'atteignent pas leurs objectifs.

Si un tel taux n'est pas inhabituel pour une technologie qui n'en est qu'à ses balbutiements, il serait dommage de ne pas tirer de leçons de ces projets.

Examinons les quatre raisons principales de l'échec des projets Big Data.

1. Objectifs trop vagues

Le motif d'échec le plus cité dans le cadre de l'enquête est l'« envergure trop peu précise » du projet. De trop nombreuses entreprises mettent en place des projets ambitieux – mais beaucoup trop ambitieux – et sans objectifs clairs, et ne s'en sortent pas lorsqu'elles doivent prendre des décisions difficiles et indiquer ce qui est important et ce qui ne l'est pas.

2. Attentes mal gérées

Suite à l'hypermédiatisation des Big Data, certaines entreprises surestiment dangereusement le potentiel de leurs projets. Aussi tentant soit-il de promettre des progrès spectaculaires en un temps record, il est important que vous restiez réalistes quant à vos attentes vis-à-vis du projet, du temps et des efforts qu'il nécessitera. Si vos attentes en matière d'impact et d'informations sont trop élevées, vous vous retrouverez à parcourir des téraoctets de données inconnues à la recherche de la perle rare. Si vos attentes en matière de livraison ne sont pas réalistes, vous vous fixerez des échéances et des budgets complètement inadaptés.

¹ InformationWeek, « [Vague Goals Seed Big Data Failures](#) ».



3. Échéances dépassées et retards

Étant donné la nouveauté de la discipline pour l'entreprise, il n'est pas étonnant que la plupart des projets Big Data dépassent le budget qui leur était alloué ou prennent plus de temps que prévu. Ce problème est généralement dû à des attentes mal gérées associées à un manque de compréhension du processus de mise en oeuvre d'une architecture évolutive. Les développeurs Hadoop-Java sont rares et coûtent cher, et quand une entreprise en engage et leur demande de se charger du codage manuel d'immenses implémentations, elle se rend bien vite compte qu'il sera impossible de quitter l'environnement de sandbox sans erreurs. Par conséquent, les projets Big Data finissent par être traités comme une expérience scientifique et ne quittent jamais la phase de test.

4. Incapacité d'évoluer

Trop souvent, les entreprises favorisent les opportunités à court terme plutôt que sur la durée. Bien que nous manquerions à notre devoir si nous suggérions que vous pouvez toujours éviter ce compromis, nous n'insisterons jamais assez sur l'importance d'une vision à long terme. Si vous voulez vous assurer que vos données sont correctement sécurisées et gérées, vous devez garder à l'esprit les implications à long terme de votre projet.

Les quatre raisons de l'échec des projets Big Data sont inquiétantes et beaucoup trop courantes. Examinons la manière dont vous pouvez les éviter et créer une implémentation durable.

Comment assurer le succès de votre projet Big Data

Si la plupart des projets Big Data n'aboutissent pas en raison d'un manque de clarté ou d'une incapacité de démontrer les aspects pratiques de l'initiative, vous devez vous efforcer de faire de votre projet une priorité et de prouver son utilité. Voici trois conseils utiles pour vous assurer que votre projet démarre sur les chapeaux de roue et tienne la distance.

1. Fixez des objectifs précis et gérez les attentes

Si vous n'êtes pas sûr de ce que vous souhaitez accomplir dans le cadre de votre projet, pensez aux objectifs que vous vous êtes fixé pour votre infrastructure de données existante.

Si votre entreprise a déjà besoin de données pour certains processus métiers (tels que la détection de la fraude ou les études de marché), réfléchissez à la manière dont les Big Data peuvent améliorer ces processus ou en augmenter la valeur. Plutôt que de vous attaquer à un nouveau problème, vous devriez essayer d'améliorer un processus ou un projet existant.

Si vous n'avez pas d'objectifs clairs et que vous ne pouvez pas démontrer la valeur de votre projet aux utilisateurs métiers, votre projet est voué à l'échec.

2. Définissez les indicateurs qui prouvent la valeur de votre projet

Définir clairement les indicateurs correspondant à vos objectifs peut considérablement vous faciliter la vie. En vous fixant des objectifs réalistes et mesurables, vous vous assurez que tout le monde pourra constater vos progrès.

Plus important encore, vos collègues sauront ce que vous prévoyez à long terme. Demandez-vous comment vous pouvez mesurer l'impact de votre projet dans le contexte de vos objectifs.

Ce point est crucial car votre projet exigera des compromis à court terme que vous devrez justifier auprès de vos utilisateurs métiers et la mise en place d'objectifs mesurables vous aidera à prouver que votre projet apporte une plus grande valeur ajoutée qu'ils ne l'imaginent.

3. Abordez les outils et le codage manuel de manière stratégique

Ne succombez pas à la tentation de coder tout manuellement directement dans Hadoop. N'oubliez pas que votre objectif n'est pas de créer une implémentation efficace de toutes pièces, mais de faire profiter de la valeur des Big Data à votre entreprise.

Plutôt que d'essayer de coder manuellement chaque intégration et chaque analyse et de nettoyer chaque ensemble de données, vous devriez rechercher des outils et des systèmes d'automatisation qui vous aideront à accélérer ces processus.

Plus important encore, ne tombez pas dans le piège de gaspiller le talent rare et coûteux des développeurs Java en leur confiant des tâches qui ne peuvent pas évoluer ou être transférées à d'autres employés. Votre rôle consiste à prendre des décisions stratégiques concernant le déploiement de ressources limitées de manière à atteindre vos objectifs.

Adoptez des outils capables d'augmenter la productivité de votre équipe de développement en exploitant les compétences et connaissances de vos experts en ETL, qualité des données et business intelligence (BI), de manière à permettre à vos développeurs Java de se concentrer sur des logiques spécifiques pour lesquelles aucun outil n'est disponible.

De plus, étant donné que les technologies telles que Hadoop évoluent tous les jours, vous pourriez mettre en place une couche d'abstraction capable de vous protéger des spécifications en constante évolution des technologies sous-jacentes.

Et surtout, n'oubliez pas que les compétences dont vous avez besoin sont limitées – mais que les outils sont toujours disponibles.

Choisir un projet adapté

Compte tenu des défis qui vous attendent, examinons comment choisir un projet adapté à votre entreprise.



Caractéristiques d'un projet adapté

Si votre entreprise a soif de changement et a déjà accepté qu'elle aura besoin d'un cadre de gouvernance de données complet pour améliorer sa façon de travailler, vous pouvez probablement vous permettre de passer cette section.

Mais si vous envisagez de mettre en place un projet tactique localisé qui pourra ensuite être adapté à l'ensemble de l'entreprise, nous vous invitons à poursuivre votre lecture.

Un projet adapté se compose des quatre éléments suivants.

1. Démonstration de la valeur

Pour qu'un projet soit adapté, sa valeur doit être partagée équitablement entre le service informatique et la division opérationnelle que vous essayez d'aider. Vous devez donc montrer clairement la valeur de votre projet à un service, une division opérationnelle ou un groupe, d'une manière parlante pour celui-ci.

2. Support

Pour que votre projet soit un succès, vous avez absolument besoin du soutien de dirigeants qui partagent votre vision. Les projets Big Data nécessitent le soutien de personnes haut placées qui sont prêtes à défendre votre travail.

Par exemple, si vous savez que vous pouvez créer un fantastique projet d'analyse pour le service logistique mais que le seul responsable qui vous soutienne est le directeur marketing, vous devez repenser votre approche. Si vous avez le soutien du directeur marketing, élaborer un projet capable de répondre aux exigences en termes d'analyse du service marketing. Vous ne pouvez obliger personne à accepter le changement. Suivez les personnes qui ont de l'influence et tirez-en le meilleur parti possible.

3. Effet quilles

L'importance stratégique de votre premier projet tactique est cruciale.

Vous devez prouver que les Big Data peuvent indubitablement aider la division opérationnelle que vous visez, mais également vous assurer que sa valeur peut ensuite être facilement communiquée au reste de l'entreprise.

Lorsque vous sélectionnez votre premier projet, faites un choix stratégique.

Une fois que vous avez prouvé la valeur des Big Data à votre service marketing, par exemple, vous éprouverez moins de difficultés à convaincre les équipes logistiques qui, dans le cas contraire, auraient pu se montrer réticentes.

4. Compétences transférables

Comme nous venons de le mentionner, il est important que la valeur de votre premier projet contribue à convaincre les autres services de votre entreprise. C'est pourquoi vous devez vous assurer que vous pouvez tirer des compétences, des capacités et des leçons de votre premier projet. Et surtout, vous devez vous assurer de les consigner afin de pouvoir les transférer à votre prochain projet. N'oubliez pas que, pour réussir, vous devez assurer le succès de vos projets à venir.

Vous devez donc être prêt à évoluer afin de traiter davantage de projets à l'avenir. Il ne suffit pas de faire évoluer votre cluster. Vous devez également faire évoluer vos compétences et vos opérations. Vous devez soit engager plus de développeurs Java/Hadoop, soit trouver le moyen de tirer meilleur parti des ressources que vous possédez déjà.

Réfléchir à son impact

Lorsque vous réfléchissez à votre prochain projet, tenez compte de l'impact qu'il aura sur votre entreprise. Basez-vous sur les trois aspects majeurs suivants pour déterminer si le projet Big Data que vous envisagez est adapté.

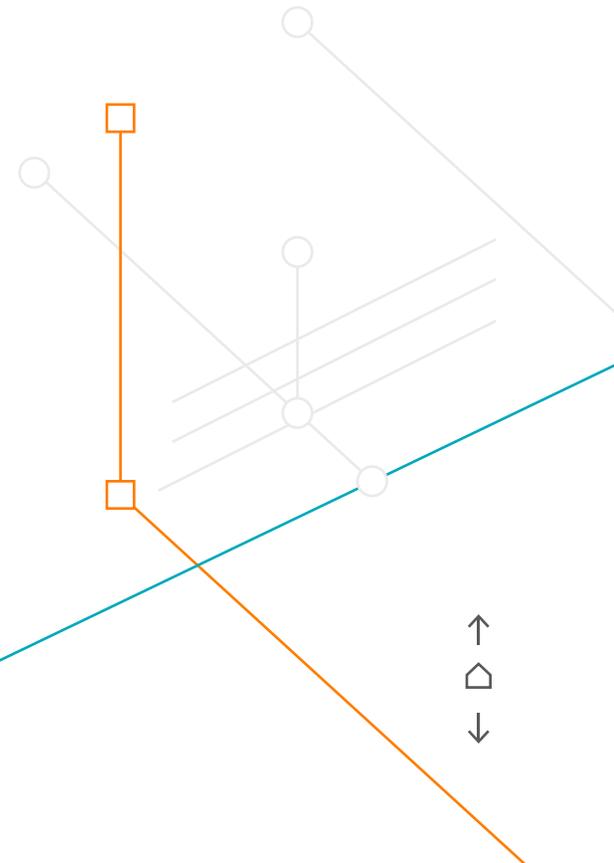
1. Coût et perturbation des activités

Au niveau le plus basique, le coût de votre projet est basé sur le temps et l'argent que vous devrez y consacrer. En réalité, vous devriez également tenir compte des perturbations qu'il pourrait entraîner.

Parfois, les perturbations concernent les procédures – lorsque les divisions opérationnelles ont l'habitude de contrôler leurs données et ne sont pas enthousiastes à l'idée de passer par un cadre de gouvernance de données centralisé.

Les perturbations peuvent également être d'ordre technologique et liées aux compétences – lorsque vous devez intégrer de nouvelles technologies dans votre infrastructure existante et réorganiser ou mettre à niveau les compétences nécessaires.

Quoi qu'il en soit, vous devez prévoir les perturbations, les reconnaître et les réduire au maximum – ou démontrer que votre projet en vaut la peine.



2. Timing des avantages et de l'impact

Lorsque vous devez choisir entre plusieurs projets de départ, vous aurez tendance à vous tourner vers ceux qui auront l'impact le plus important sur l'entreprise et l'amélioreront au maximum. Mais il est également important de tenir compte de la nature de l'impact sur l'entreprise. L'entreprise bénéficiera-t-elle de la majeure partie de la valeur à court terme ou à long terme ?

Plus important encore, quand les utilisateurs métiers constateront-ils l'impact du projet sur l'entreprise ? Par exemple, vous pourriez introduire la gestion des données de référence dans votre data warehouse et améliorer considérablement l'efficacité de votre business intelligence. Mais la valeur de ce projet ne se fera ressentir que quand vos analystes métiers se rendront compte qu'ils ne devront plus jamais nettoyer de données financières.

3. Ressources et restrictions

Compte tenu de l'analyse des deux facteurs ci-dessus, prenez en considération les ressources à votre disposition. Nous aborderons ce sujet plus en détail ultérieurement. En attendant, gardez à l'esprit qu'il est naturel de vouloir que votre projet vous rapporte plus que ce que vous y investissez.

Mais pour atteindre cet objectif, vous devez être prêt à faire un compromis. D'un côté, vous visez un impact maximal sur votre entreprise. Mais vous devez également choisir de manière stratégique la manière dont vous allez dépenser votre budget. Vous rêveriez certainement de mettre en place une équipe de spécialistes des données digne de Google, mais pouvez-vous vraiment vous le permettre ? Pour assurer le succès de votre projet, vous devez absolument faire les bons choix en matière d'outils et de personnel.

Quelques exemples de projets Big Data tactiques

Une vaste gamme d'applications se prêtent aux Big Data. Aussi palpitante soit-elle, cette technologie peut également être un peu intimidante pour les entreprises qui ne savent pas par quel projet commencer. Voici une liste de projets Big Data tactiques que nos clients ont entrepris.

Si vous ne savez toujours pas par quel projet commencer, jetez un coup d'oeil aux exemples suivants pour découvrir ce que les Big Data peuvent apporter à votre entreprise.

Finance

- Analyse des risques et des portefeuilles
- Recommandations en matière d'investissements

Distribution

- Engagement proactif des clients
- Services basés sur la localisation

Secteur de la santé

- Prédications quant au sort des patients
- Coût total des soins
- Recherche de nouveaux médicaments

Industrie

- Logiciels embarqués pour voitures
- Maintenance prédictive

Secteur public

- Assurance maladie
- Échanges
- Optimisation fiscale
- Détection de la fraude

2^e partie

Votre stratégie

Nous allons maintenant passer à la pratique et analyser les exigences spécifiques de votre prochain projet Big Data.

Définir vos objectifs

Prenez un crayon. Comme nous l'avons déjà mentionné, la première cause d'échec des projets Big Data est le manque d'objectifs clairement définis. Assurons-nous maintenant que le projet auquel vous pensez n'est pas ambigu.



Objectifs informatiques

Examinons maintenant les objectifs informatiques en rapport avec votre projet.

(N'oubliez pas que si votre projet vise à aider le service informatique à travailler plus efficacement ou plus rapidement, vous aurez du mal à convaincre les utilisateurs métiers. C'est pourquoi les objectifs informatiques doivent être mentionnés en même temps que les objectifs qui intéressent vos utilisateurs métiers.)

Indiquez par ordre d'importance le ou les objectifs de votre projet Big Data en rapport avec le service informatique (n'hésitez pas à ajouter un nombre inférieur ou supérieur d'objectifs).

Ex. : mettre en place des processus de collecte en temps réel, de nettoyage, de maîtrise et de stockage des données clients agrégées, des données d'utilisation de carte de crédit, des données de graphes sociaux et des indicateurs de perte de clients

Arrêtez-vous, collaborez et écoutez

Nous avons créé ce manuel pour vous aider à démarrer votre projet Big Data, que vous travailliez du côté des utilisateurs métiers ou du service informatique. Dans les deux cas, ne laissez pas vos objectifs au hasard. Si vous avez besoin de conseils spécifiques pour vous aider à définir vos objectifs, collaborez avec un partenaire disposant des compétences nécessaires.

La collaboration stratégique joue un rôle primordial dans la réussite de votre projet.

Objectifs informatiques

Indiquez une durée minimum et une durée maximum pour chaque objectif.

Ex. : deux à quatre mois

Ensuite, pour chaque objectif, indiquez une mesure de succès que vous pouvez utiliser pour déterminer si l'objectif a été atteint. Dans l'idéal, il doit s'agir d'indicateurs disponibles ou de calculs réalisés à partir de ces indicateurs.

Ex. : taux de prédictions exactes en matière de perte de clients de X %

Définir vos besoins en données

Maintenant que nous avons mis en évidence les objectifs spécifiques de votre projet Big Data, entrons dans le vif du sujet – les données elles-mêmes. Quel que soit votre projet, vous devez réfléchir de manière stratégique aux informations dont vous avez besoin, aux ensembles de données qui répondent à ce besoin, à la manière dont vous allez collecter les données et à la façon dont vous allez les utiliser.

De quelles données avez-vous besoin ?

Penchons-nous tout d'abord sur l'objectif le plus basique de votre projet Big Data : les informations que vous souhaitez fournir à votre entreprise. Répondez aux questions suivantes le plus précisément possible.

Pour atteindre les objectifs métiers mentionnés précédemment, de quelles informations mes utilisateurs métiers disent avoir besoin pour prendre des décisions en toute connaissance de cause ?

Ex. : parmi vos meilleurs clients, lesquels êtes-vous susceptible de perdre et quels comportements correspondent à la perte de clients ?

Quelles données peuvent être utilisées pour fournir ces renseignements ?

Ex. : historique des achats des clients, avis, taux d'achat, taux d'abandon, taux de rebond, qualité du service aux clients

Quels ensembles de données auxquels je n'ai actuellement pas accès pourraient contenir des renseignements contextuels supplémentaires ?

Ex. : données sociales tierces, données de marché, données météorologiques

La chasse aux données dont vous ignorez l'emplacement

Lorsque vous réfléchissez aux ensembles de données auxquels vous n'avez pas accès, ne vous limitez pas aux données externes à votre entreprise. Gartner a constaté que la plupart des entreprises n'utilisaient que 15 % de leurs données. Appfluent, une société qui se consacre à l'analyse statistique de l'utilisation des data warehouses, a remarqué qu'entre 30 et 70 % des données d'un data warehouse étaient en sommeil. Le reste des données est caché dans des silos, des archives héritées et des magasins de données difficiles d'accès, coûteux ou difficiles à trouver. Cela ne serait pas un problème si ce n'est que vous payez déjà pour stocker toutes ces données.

Lorsque vous recherchez les données dont vous avez besoin, nous vous conseillons d'examiner les données dont votre entreprise dispose déjà.



Cinq considérations clés en matière de données

Une fois que vous avez identifié les données que vous allez rechercher, vous aurez une vue plus claire des défis qui vous attendent en matière de Big Data.

Il y a cinq éléments clés que vous devriez prendre en considération avant d'aller plus loin, car ils dicteront les besoins pour chaque ensemble de données, ainsi que pour votre ensemble de données Big Data.

1. Préparez-vous pour le volume

Vous allez devoir vous préparer à traiter le volume important de données dont vous avez besoin. Classez vos données en fonction de leur valeur (transactions clients, par exemple), de leur utilisation (fréquence d'accès), de leur taille (gigaoctets, téraoctets), de leur complexité (données générées par des systèmes automatisés, données relationnelles, vidéos...) et des personnes autorisées à y accéder (vos spécialistes des données ou tout utilisateur métier).

Un inventaire complet et organisé vous aidera à déterminer comment gérer toutes ces données. Évaluez vos capacités de stockage et de traitement actuelles et recherchez les moyens les plus rentables et efficaces de les rendre évolutifs.

2. Tenez compte de la variété

L'aspect le plus complexe des Big Data réside dans la multitude de formats et de structures que vous allez devoir rapprocher dans vos analyses. Vous allez devoir intégrer de nombreuses sources si vous souhaitez inclure de nouveaux types et structures de données (données sociales, données issues de capteurs, données vidéo) aux sources auxquelles vous êtes déjà habitué (bases de données relationnelles, mainframes hérités).

Le codage manuel de chaque intégration est si fastidieux qu'il nécessiterait tout le temps et toutes les ressources dont vous disposez. Tirez le meilleur parti des outils d'intégration de données et de qualité des données disponibles pour accélérer vos processus pour les tâches les plus importantes.

3. Gérez la vitesse

La combinaison de données en streaming en temps réel et de données historiques améliore généralement les capacités prédictives de l'analyse. Il est donc possible que certaines des données dont vous avez besoin ne soient utiles que si elles alimentent constamment vos systèmes.

En effet, la plupart des analyses en temps réel doivent se baser sur les flux de données – souvent issus de différentes sources, et dans différents formats. Préparez votre projet en prévoyant une technologie d'analyse du streaming ainsi qu'une infrastructure logique pour gérer toutes ces données.

4. Assurez-vous de la véracité de vos données

Aussi importantes vos analyses soient-elles, elles n'auront aucune valeur si les utilisateurs ne peuvent pas faire confiance aux données sur lesquelles elles sont basées. Plus vous analysez de données, plus il est important que vous mainteniez un niveau élevé de qualité des données.

Pour que vos données soient adaptées à leur utilisation, vous devez d'abord définir l'utilisation que vous en ferez. Si un spécialiste des données cherche à dégager des tendances dans les données clients agrégées, la préparation nécessaire est minime. D'un autre côté, les données des rapports financiers et de la chaîne logistique devront être hautement organisées, nettoyées et certifiées pour garantir leur exactitude et leur conformité.

Créez des catégories basées sur la quantité de préparation nécessaire, des données brutes à un magasin de données de référence hautement organisé comprenant des données nettoyées, fiables et faisant autorité.

5. Pensez à la conformité

Les différents ensembles de données s'accompagneront de normes et d'exigences de sécurité différentes. Pour chaque ensemble de données, vous devez réfléchir à une façon de rendre les données anonymes conformément aux stratégies de sécurité.

D'énormes quantités de données proliféreront dans des centaines de magasins de données au sein de votre entreprise.

Vous devez savoir où se trouvent vos données sensibles, vous assurer de les sécuriser à la source grâce au cryptage et contrôler leur accès.

Même lorsque vos données sensibles sont stockées dans un système d'archivage intelligent et sécurisé, masquez-les à l'aide de règles prédéfinies chaque fois qu'elles migrent ou qu'elles sont transférées dans vos environnements de développement et de test.

En appliquant ces cinq considérations à chacun de vos ensembles de données, vous serez prêt à relever le défi que représentent les Big Data de manière plus réaliste.



3^e partie

Votre chaîne logistique Big Data lean

Les méthodes de business intelligence et de data warehouse traditionnelles ne sont pas en mesure d'évoluer pour répondre aux besoins des initiatives Big Data. Nous allons maintenant examiner comment vous pouvez faire évoluer votre équipe, vos processus et votre infrastructure.

Votre équipe

Votre équipe chargée des Big Data représente à la fois votre plus grand défi et votre plus grande opportunité. Vous devez trouver le juste milieu entre des personnes qui comprennent les objectifs métiers et des personnes capables de mettre en oeuvre vos exigences techniques.



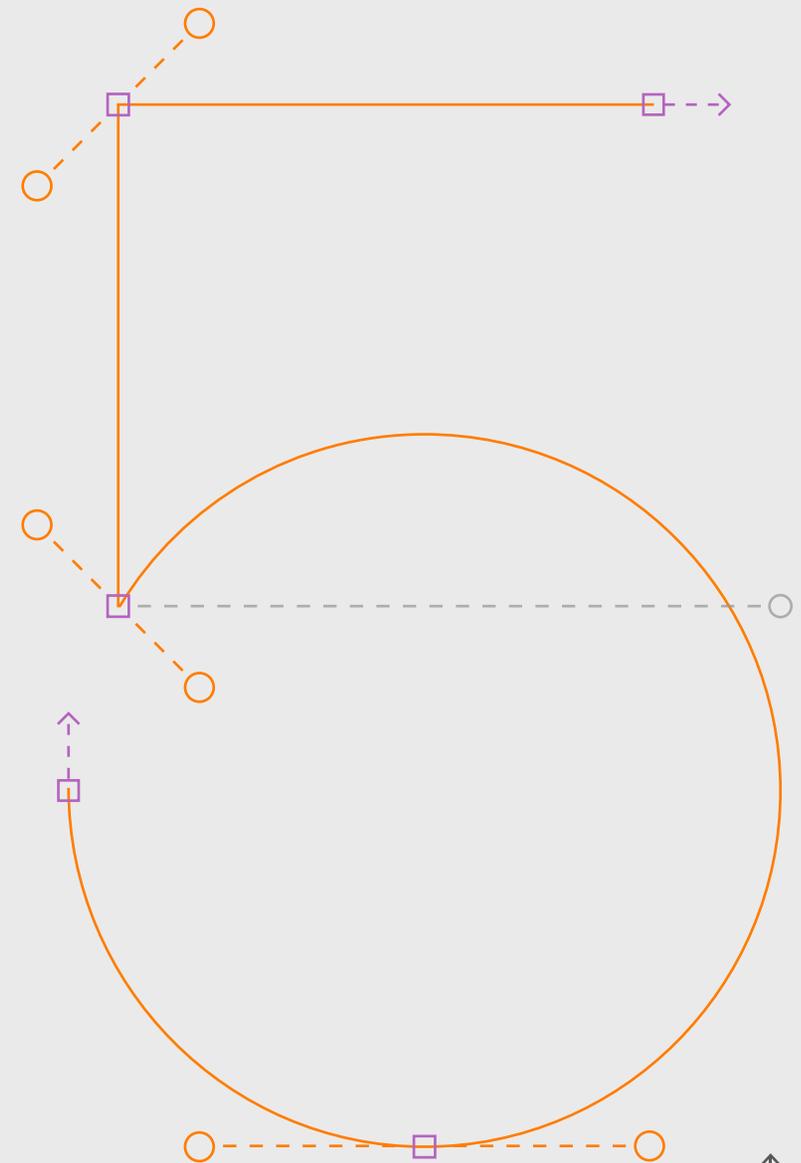
Cinq leçons fondamentales pour renforcer l'esprit d'équipe

La plupart des entreprises sous-estiment le niveau de compétences nécessaire pour appliquer efficacement une nouvelle technologie telle que Hadoop.

Les infrastructures de données distribuées sont très difficiles à gérer. Des compétences Java nécessaires pour développer sur Hadoop aux nouvelles compétences liées aux données pour lesquelles vous devrez engager, vous allez devoir apporter de nombreuses nouvelles compétences à votre projet pour assurer son succès.²

Quand vous commencez à créer votre équipe, assurez-vous d'intégrer les leçons suivantes à votre stratégie de recrutement.

² « [Hadoop, Python, and NoSQL lead the pack for big data jobs](#) », InfoWorld, 5 mai 2014



1. Utilisez les compétences pour lesquelles vous avez engagé vos employés

L'une des plus grandes erreurs des entreprises lorsqu'elles engagent des spécialistes des données et des analystes quantitatifs consiste à leur confier les tâches ingrates. Lorsque vos employés les plus qualifiés passent leur temps à intégrer manuellement et à nettoyer des données, non seulement ils sont frustrés – mais en plus, vous passez à côté de l'occasion de tirer parti des compétences que vous avez eu tant de mal à trouver.

Assurez-vous de confier aux personnes dotées de compétences rares des tâches qui les nécessitent vraiment. Vous ne voudriez pas que vos travailleurs les plus qualifiés quittent votre entreprise et il serait vraiment dommage qu'ils perdent leur temps à accomplir des tâches que vous pourriez facilement automatiser.

2. Réfléchissez de manière stratégique à la composition de votre équipe

Si tout se passe bien, votre projet évoluera aussi bien en matière d'envergure que de ressources. Réfléchissez de manière stratégique dès maintenant pour éviter de vous rendre compte trop tard que vous ne pouvez pas faire évoluer certains processus suffisamment rapidement car vous disposez d'un nombre limité d'employés disposant des compétences nécessaires – même si vous vous trouvez dans la Silicon Valley.

Si l'envergure de votre projet augmente, quelles compétences êtes-vous sûr de pouvoir trouver à temps afin de répondre à vos besoins ? Par exemple, les spécialistes des données sont beaucoup plus difficiles à trouver, à former et à engager que les développeurs.³

Il est crucial d'atteindre un équilibre au sein de votre équipe. Vous devez trouver un équilibre entre l'expérience dans le domaine de la gestion de données et l'enthousiasme nécessaire pour apprendre à utiliser de nouveaux outils. De plus, vous devez trouver le juste milieu entre les personnes dotées des compétences techniques recherchées et les experts du domaine dotés des connaissances nécessaires pour créer les modèles adaptés.

3. Alignez les objectifs de votre projet dès le départ, et communiquez-les

L'une des erreurs que les entreprises font le plus souvent quand elles engagent de nouveaux employés consiste à oublier de communiquer les véritables objectifs du projet. Du premier entretien au premier jour de travail, vous devez indiquer clairement ce que vous essayez d'apporter aux utilisateurs métiers. Exploitez le soutien dont vous disposez au sein de la direction pour communiquer votre mission et partager vos réussites ainsi que les problèmes que vous rencontrez.

Si vos nouvelles recrues ne saisissent pas bien la valeur métier de votre projet, elles risquent de penser qu'elles doivent uniquement se concentrer sur les objectifs du service informatique.

³ « [Big Data's High-Priests of Algorithms](#) », Wall Street Journal, 8 août 2014

4. La nécessité de gérer votre équipe augmente au fur et à mesure qu'elle s'agrandit.

Contrairement aux nouvelles technologies qui peuvent être déployées, implémentées et intégrées de manière objective, les nouveaux employés doivent s'habituer à leur nouvel environnement de travail, à leurs nouvelles tâches et à leurs nouveaux objectifs. Que ce soit vous ou l'un de vos collègues, quelqu'un doit relever le défi en termes de gestion lié à la mise en place d'une nouvelle équipe.

Ne sous-estimez pas les éléments tels que la culture et l'esprit d'équipe. Réfléchissez bien à la manière dont vous intégrez les nouvelles recrues à vos processus. Vous ne pourrez peut-être pas les former en termes de compétences, mais vous pouvez certainement les aider à devenir de meilleurs membres d'équipe.

5. Votre équipe ne peut pas se permettre de stagner

Des technologies Big Data voient le jour quotidiennement. Et celles qui existent déjà évoluent rapidement. Nous vivons à une période palpitante pour les entreprises suffisamment courageuses pour adopter des meilleures pratiques très tôt. Mais le fait d'avoir une longueur d'avance sur la concurrence s'accompagne d'un défi considérable.

Les compétences de vos employés doivent évoluer aussi rapidement que le monde qui les entoure. La bonne nouvelle, c'est que rien n'est plus motivant que le défi consistant à garder une longueur d'avance. Le plus difficile est de parvenir à fournir les formations et les discussions nécessaires pour développer leurs capacités et les vôtres.

L'importance d'une approche stratégique

Vous devrez continuellement décider si vous souhaitez développer vos capacités à l'aide d'outils automatisés ou d'intégrations manuelles.

Le codage manuel vous offre un contrôle précis et total de ce que vous créez. Ce contrôle est souvent inestimable et il peut être nécessaire, par exemple si vous écrivez un script complexe visant à extraire les métadonnées d'une façon inédite.

Quant aux outils, ils vous offrent une plus grande flexibilité et la capacité de répéter le même processus de manière durable. Ceci est crucial pour des tâches telles que l'intégration de données et la qualité des données, car cela signifie que vos analystes et spécialistes des données super intelligents ne sont pas obligés de s'occuper des tâches ingrates.

Faites preuve de réalisme en ce qui concerne vos ressources. Si vous ne parvenez pas à mettre en place une équipe aussi imposante et brillante que celle de Google, ne gaspillez pas le peu de ressources dont vous disposez à essayer.



Implémenter la gouvernance de données

Si vous développez un projet Big Data plus fondamental (et nous espérons que ce ne sera qu'une question de temps), vous devrez mettre en place un cadre procédural pour la gouvernance de données. D'ailleurs, même si votre projet Big Data ne vise à apporter de la valeur qu'à un seul service, vous pourriez envisager de créer un mini-comité de gouvernance de données afin d'apprendre à faire face aux défis uniques que représente un tel organisme.

Votre comité de gouvernance de données correspond essentiellement à l'organisme officiel composé de dirigeants visant à superviser l'approche de l'entreprise vis-à-vis des données. Mais votre projet nécessite également des gestionnaires de données – des personnes d'une fonction ou d'un service spécifique chargées de gérer les données d'une division opérationnelle particulière.

(En fait, certains de nos clients choisissent leurs gestionnaires de données en fonction du domaine de données, ce qui signifie qu'une personne peut être chargée des données produits, tandis qu'une autre s'occupe des données clients, et ainsi de suite.)

Votre objectif doit être de créer des processus visant à garantir que votre cadre de gouvernance de données offre plus d'avantages que d'inconvénients. Vous devez travailler de manière proactive pour vous assurer qu'il ne se transforme pas en fardeau bureaucratique en vous assurant que tous les employés cherchent à atteindre les mêmes objectifs dans les mêmes délais.

Dans l'idéal, votre cadre de gouvernance de données devrait présenter les cinq caractéristiques suivantes.

1. Interfonctionnalité

Un comité de gouvernance de données composé de personnes dont le rôle est similaire s'avérera inefficace. Vous devez créer un organisme qui représente les vues et les besoins uniques de chaque division opérationnelle concernée par votre projet Big Data.

2. Communication

Sans une communication efficace entre les fonctions, services et domaines, votre projet sera submergé par la bureaucratie et les malentendus. Ce phénomène est beaucoup trop courant. Assurez-vous d'apaiser toutes les craintes ou de les aborder de manière adaptée.

3. Efficacité

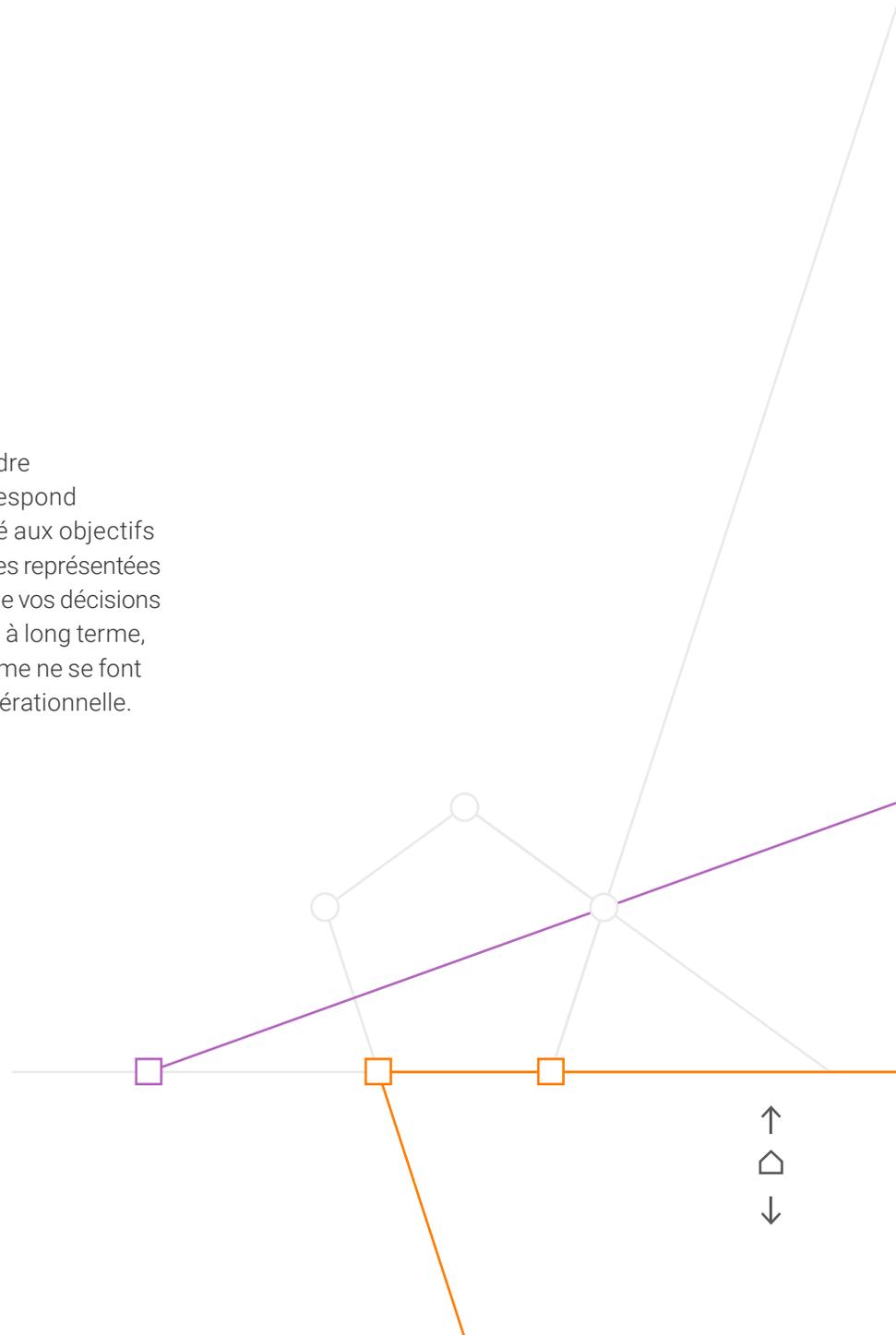
Vous ne devez pas voir votre processus interfonctionnel comme un obstacle. Vous devez vous montrer très flexible pour assurer le succès de votre projet Big Data. Nous vous conseillons donc d'intégrer autant de règles d'automatisation et de création de rapports sur les exceptions que possible et d'adopter des outils de collaboration afin de garder les canaux de communication ouverts et rapidement accessibles.

4. Validation

Communiquez efficacement les objectifs principaux de votre projet et assurez-vous que chaque personne impliquée dans votre cadre de gouvernance de données met tout en oeuvre pour les atteindre. Votre réflexion et votre prise de décision dans le domaine de la gouvernance doivent être guidées par des objectifs communs.

5. Centralisation

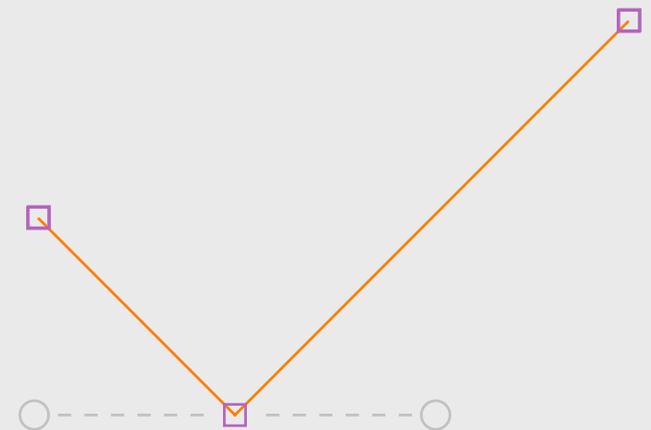
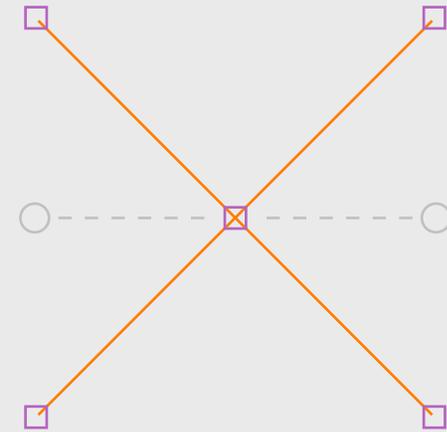
Le plus grand défi posé par un cadre de gouvernance de données correspond à la nécessité de donner la priorité aux objectifs de l'une des divisions opérationnelles représentées au sein du comité. Assurez-vous que vos décisions profiteront à l'ensemble du comité à long terme, même si les avantages à court terme ne se font ressentir que dans une division opérationnelle.



Les compétences dont vous avez besoin et celles dont vous disposez déjà

C'est le moment de ressortir votre crayon. Maintenant que vous connaissez les différents pièges et opportunités subjectifs associés à la création de votre nouvelle équipe, déterminons les différents membres qui la composeront.

La page suivante répertorie les rôles liés aux Big Data basés sur les fonctions pour lesquelles nos clients ont recruté. Selon le personnel déjà à votre disposition et la durée prévue de votre projet ([telle qu'indiquée dans la section commençant à la page 24](#)), indiquez le nombre de personnes que vous avez besoin de recruter.



Le rôle	L'un de mes employés peut-il déjà remplir ce rôle ?	Je dois recruter pour ce rôle	Étant donné le temps à ma disposition, je dois engager X personnes
Spécialiste des données			
Expert de domaine			
Analyste métier			
Analyste de données			
Ingénieur de données			
Administrateur de base de données			
Architecte d'entreprise			
Architecte de solutions d'entreprise			
Architecte de données			
Gestionnaire de données			
Développeur ETL (intégration de données)			
Développeur d'applications			
Développeur de tableaux de bord			
Créateur de modèles statistiques			
Autre			

Le besoin d'une réflexion intégrée

Quand vous recherchez de nouveaux membres d'équipe, ne vous limitez pas aux personnes qui possèdent les qualifications adaptées. Ne vous y trompez pas – trouver des personnes disposant des qualifications adéquates représente déjà un défi en soi. Mais vous devez également rechercher des personnes désireuses de combiner objectifs métiers et capacités techniques.

Nos clients ne cessent de souligner combien il est important que les personnes qui rejoignent leurs projets Big Data soient capables de comprendre la réalité du monde des affaires et d'effectuer des tâches scientifiques liées à des données complexes. Une telle réflexion intégrée est aussi cruciale que rare. Elle vaut la peine d'y former votre équipe et de la récompenser.

Comprendre les outils liés aux Big Data

Passez en revue cette liste d'outils et placez 1 X devant ceux qui sont les plus importants – et les plus pertinents sur le plan stratégique – dans le cadre de votre projet spécifique.

Analyse

Outils et processus qui transforment les données brutes en renseignements, tendances, prédictions et calculs au sujet du domaine que vous analysez.

Visualisation

Êtes-vous en mesure de présenter vos données et vos résultats d'une manière facile à assimiler et à comprendre ?

Analyses avancées

Êtes-vous en mesure d'appliquer des algorithmes d'analyse de pointe à vos ensembles de données afin d'effectuer des calculs complexes ?

Apprentissage machine

Êtes-vous en mesure d'appliquer des algorithmes d'apprentissage machine pour identifier des tendances et effectuer des prédictions à un niveau pour lequel vous ne disposez pas de la bande passante nécessaire ?

Parmi ces outils et technologies, certains outils tels que l'intégration de données, la qualité des données et la gestion des données de référence sont si fondamentaux pour votre projet Big Data qu'ils ne valent pas la peine d'être recréés entièrement. Le temps et les ressources qui seraient nécessaires pour que vous mettiez en place ces capacités vous-mêmes ne valent pas la peine que vous y consacriez les précieuses compétences et le temps des personnes de votre projet Big Data.

Gardez à l'esprit les objectifs de votre projet : ceux-ci ne comprennent pas la création sur mesure de tous ses éléments.



Vos processus

Passons aux processus dont vous aurez besoin pour aborder les Big Data. Vos processus seront spécifiques à vos objectifs et exigences, mais cette section devrait vous donner un aperçu de ce à quoi vous pouvez vous attendre et de ce que vous apprendrez.



Les huit étapes vers les Big Data

Nous savons par expérience que les méthodologies flexibles sont une excellente approche des projets Big Data. Elles vous permettent de gérer les attentes, d'apprendre de vos erreurs et de parvenir aux meilleurs processus. Ceci dit, votre approche de votre projet dépend entièrement de vous et de votre situation.

Dans tous les cas, les huit étapes suivantes seront cruciales à votre chaîne logistique Big Data. Quelle que soit votre manière de procéder, assurez-vous que vous et votre équipe établissez des processus efficaces pour ces étapes.

Passez en revue cette liste d'outils et placez 1 X devant ceux qui sont les plus importants – et les plus pertinents sur le plan stratégique – dans le cadre de votre projet spécifique.

1. Accès aux données

Votre premier défi consiste à acquérir toutes les données dont vous avez besoin. Dans certains cas, cela impliquera la capture de données en streaming et dans d'autres, l'extraction de données d'une base de données. Créez des processus réutilisables et gérables pour vous assurer que ces données peuvent être stockées en accord avec la manière dont vous allez les utiliser.

2. Intégration des données

Le défi le plus complexe des Big Data est lié à la variété des structures et formats de données. Pour réaliser vos analyses de manière durable, vous aurez besoin de mettre en place un processus d'intégration et de normalisation de toutes ces données. Dans l'idéal, celui-ci devrait impliquer un minimum de traitement manuel.

3. Nettoyage des données

Pour que vos analyses soient fiables, vous devez nettoyer vos données de manière à supprimer les doublons, les erreurs, les inexactitudes et les données incomplètes. Votre processus doit permettre à vos analystes et spécialistes des données les plus qualifiés de ne pas perdre tout leur temps à « faire le ménage ».

4. Maîtrise des données

L'un des moyens de conserver une source fiable de données propres et intégrées consiste à mettre en place un processus vous permettant de maîtriser vos données. L'objectif est de créer une vaste collection de données consolidées, organisées par domaine (tel que les produits, les clients, etc.) et enrichies avec des analyses des Big Data, capables d'alimenter tous vos autres systèmes.

5. Sécurisation des données

Vous allez maintenant établir deux processus de base. Le premier consistera à définir les règles et pratiques de sécurité nécessaires pour chaque ensemble de données. Le deuxième permettra d'identifier les données sensibles et de les masquer de manière persistante et dynamique afin de vous assurer que ces règles et meilleures pratiques sont appliquées de manière cohérente.

6. Analyse des données

Votre processus d'analyse dépendra de vos analystes, de vos outils d'analyse et de vos exigences en fonction de vos objectifs. La découverte itérative et l'amélioration continue joueront un rôle crucial à cette étape, car il est important que ce processus s'améliore, s'accélère et devienne moins coûteux et plus évolutif au fil du temps et au fur et à mesure que vous acquérez de l'expérience.

7. Analyse de vos besoins métiers

Cette étape est cruciale et pourtant presque toujours négligée. Mettez en place un processus clair d'analyse des besoins métiers dès l'étape d'analyse de vos données. Ceci est essentiel car si vous cessez d'être à l'écoute des besoins métiers, vous risquez d'isoler vos efforts et de limiter votre impact sur l'entreprise.

8. Utilisation des informations

Comme nous l'avons déjà mentionné dans ce manuel, l'impact de votre projet Big Data sur votre entreprise doit se faire ressentir. Créez des pipelines automatisés afin de fournir les réponses que vous trouvez aux utilisateurs métiers qui en ont le plus besoin. Par exemple, les données concernant les clients les plus susceptibles de passer à la concurrence doivent être mises à la disposition des agents du service client par le biais d'un tableau de bord. Pensez à inclure un cycle de commentaires afin de savoir ce que les utilisateurs pensent des données.

L'importance de la consignation

Quand vous maîtriserez ces huit étapes, votre projet Big Data sera sur la bonne voie. L'important est d'établir des processus clairs, réutilisables, évolutifs et qui s'améliorent constamment. Pour ce faire, la consignation de vos processus et les améliorations qui en découlent sont cruciales pour votre équipe.

Les compétences, les capacités et les leçons tirées de votre projet Big Data doivent être transférables et vous devez les communiquer fréquemment.

Votre architecture

Pour que votre chaîne logistique Big Data soit lean et efficace, vous devez vous assurer que son architecture est bien établie et conçue de manière stratégique. Dans cette section, nous allons discuter de l'architecture Big Data idéale et de la manière de la déployer en adoptant une approche progressive.



Première étape : votre sandbox

Lorsque vous commencez à créer l'architecture de votre projet Big Data, le point de départ le plus logique est de mettre en place un environnement de développement sandbox vous permettant d'utiliser des données de test pour vous assurer de la faisabilité de votre architecture. Dans ce contexte, tenez compte des conseils suivants.

Commencez petit

En commençant dans un sandbox bien défini que vous contrôlez entièrement, vous pourrez progresser vers une implémentation des plus réussies. Démarrez le plus rapidement possible et consignez les leçons tirées de chaque itération.

La taille compte

La différence principale entre le sandbox et votre véritable implémentation est que votre environnement de production sera beaucoup plus grand. Vous aurez besoin de processus automatisés pour ingérer, intégrer, nettoyer et distribuer sa sortie. Vous aurez donc besoin d'une structure beaucoup plus performante et de composants et processus éprouvés pour assurer la fiabilité et la flexibilité de votre environnement de production réel.

Masquez avant de tester

Lorsque les entreprises utilisent des données de test, elles utilisent généralement une variante de leurs données de production réelles afin de s'assurer que les formats et structures correspondent à ceux de l'environnement de production. Malheureusement, si vous ne masquez pas correctement les données utilisées, vous risquez de laisser des données sensibles à la vue de tous dans un environnement de test extrêmement risqué.

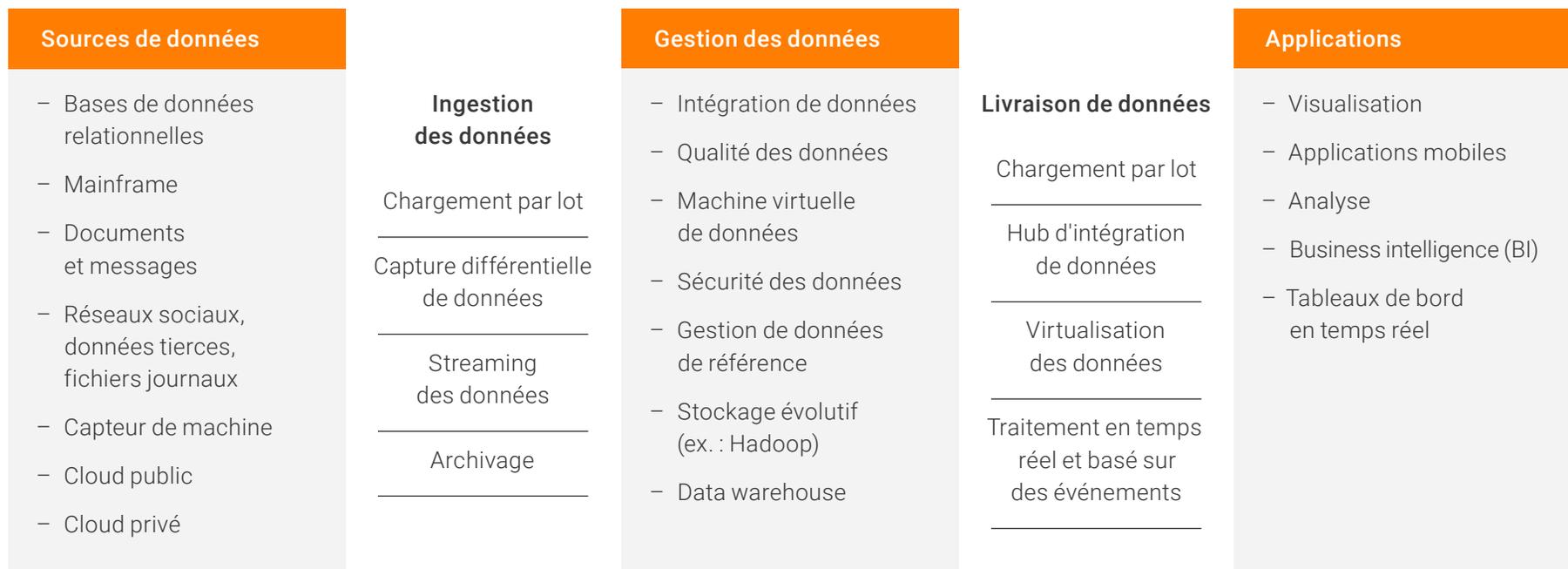
Évitez les erreurs de codage manuel

Le dépassement des échéances et les retards coûteux dans les projets Big Data sont principalement dus au fait que des erreurs de codage manuel qui n'ont pas été identifiées dans le sandbox reviennent hanter votre équipe lorsque l'architecture passe en production. Ainsi, si vous codez manuellement une grande partie de votre architecture, attendez-vous à devoir altérer une grande partie du code pour répondre aux exigences de l'environnement de production et à devoir gérer les attentes en conséquence. Vous pouvez également utiliser des outils de productivité et d'automatisation pour éviter de devoir altérer votre code et résoudre les erreurs.



L'architecture Big Data idéale

Le schéma suivant représente nos recommandations pour créer la technologie Big Data et l'architecture de processus idéales.



Votre plan de projet

Nous avons maintenant analysé chaque aspect de votre passage aux Big Data. L'étape suivante consiste à utiliser ce plan de projet pour vous aider à gérer votre projet Big Data de sa conception à son implémentation.



Votre plan de projet

Utilisez ce plan de projet pour vous aider à consigner les détails et les différents éléments de votre projet Big Data. Utilisez ensuite le document que vous avez compilé pour obtenir le soutien dont vous avez besoin de la part du reste de votre entreprise. Il vous sera également utile lorsque vous discuterez avec des partenaires externes

Étape 1 : la stratégie

Identification des objectifs métiers et informatiques

Définition des mesures du succès

Étape 2 : les données

Identification des informations dont vous avez besoin

Identification des données et sources nécessaires pour les fournir

Étape 3 : la chaîne logistique

Le personnel

- Évaluation des compétences dont vous avez besoin
- Évaluation des compétences dont vous disposez déjà

Le processus

- Accès aux données
- Intégration des données
- Nettoyage des données
- Maîtrise des données
- Sécurisation des données
- Analyse des données
- Analyse des besoins métiers

Les outils

- Informatique distribuée (ex. : Hadoop)
- Qualité des données
- Intégration de données
- Gestion de données de référence
- Masquage de données
- Visualisation
- Analyse du streaming
- Analyse
- Apprentissage machine

Étape 4 : utilisation des informations

Développement de tableaux de bord

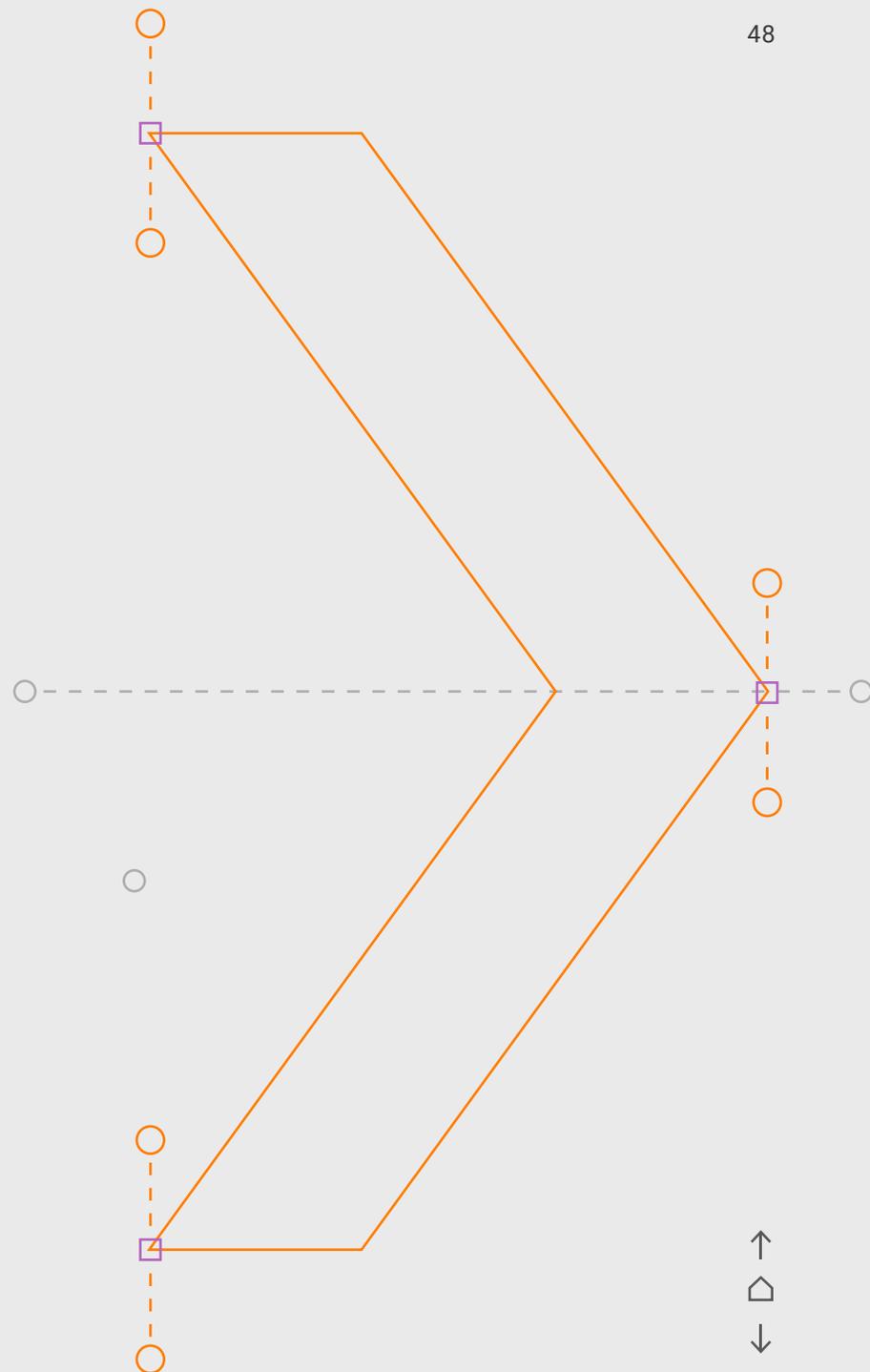
Automatisation des processus de livraison de données

Mise en place d'un processus de commentaires

Démarrer

Utilisez les listes de vérification, les principes et les directives mentionnés dans ce manuel pour apporter le potentiel des Big Data à votre entreprise. Quelle que soit la taille de votre projet, nous sommes convaincus que vous êtes désormais mieux équipé pour relever les nombreux défis qui se présenteront à vous.

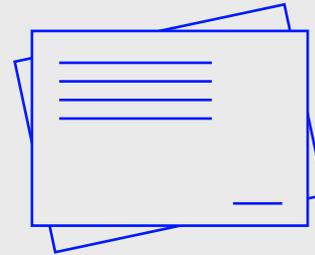
Les Big Data n'attendent plus que vous.



Étapes suivantes

Êtes-vous prêt à appliquer ce que vous avez appris ?

Si vous êtes un développeur Informatica, vous pouvez également devenir un développeur Hadoop. Nos essais de logiciels de Big Data, connecteurs et services vous guideront dans la bonne direction.



**TÉLÉCHARGER
MAINTENANT**

À propos d'Informatica

La transformation digitale fait évoluer les attentes : Les données sont la clé de la réussite des entreprises qui se doivent d'évoluer pour rester compétitives.

En tant que leader mondial dans la gestion des données Cloud d'entreprise, nous sommes prêts à vous guider de manière intelligente. Les entreprises ont besoin de solutions de données pour le Cloud, les projets Big Data, le temps réel et les flux de données en continu. Informatica est le premier fournisseur mondial de solutions de gestion de données, que ce soit dans le Cloud, sur site ou dans les environnements hybrides.

Plus de 7 000 entreprises du monde entier font appel aux solutions de données d'Informatica.

Siège mondial

2100 Seaport Blvd, Redwood City, CA 94063, États-Unis

Téléphone : +33 1 42 04 89 00 (France)

Fax : +33 1 42 04 89 01 (France)

informatica.com/fr

[linkedin.com/company/informatica](https://www.linkedin.com/company/informatica)

twitter.com/InformaticaFr

NOUS CONTACTER

