

POURQUOI LA RÉUSSITE D'UNE
STRATÉGIE CLOUD HYBRIDE
COMMENCE PAR UNE STRATÉGIE DATA

5 points clés pour réussir sa stratégie cloud hybride



RÉUSSITE
D'UN
CLOUD
HYBRIDE

Table des matières

Introduction	3
Pourquoi adopter une stratégie données d'entreprise ?	3
Importance d'une architecture de données moderne	4
1. Utiliser le cloud comme une infrastructure et non comme une architecture	4
2. Prioriser l'open source	4
3. Trouver l'équilibre entre les besoins de l'entreprise et ceux des équipes informatiques	4
4. Assurer la sécurité et la conformité des données sensibles	5
5. Gérer les coûts et optimiser les ressources	5
Étapes suivantes	9

La stratégie de données est importante,

et elle devrait vous permettre de répondre aux questions suivantes sur vos données, quel que soit leur emplacement, et qu'elles soient distribuées sur le cloud ou dans des lacs de données sur site :

- Quelles données détenez-vous ?
- Où résident-elles ?
- Comment les administrez-vous et les sécurisez-vous ?
- Comment y accédez-vous ?
- De quelle façon les valorisez-vous ?

Introduction

Les entreprises d'aujourd'hui passent au cloud à des vitesses inédites. Neuf sociétés sur dix auront au moins une partie de leurs applications ou de leur infrastructure dans le cloud d'ici à 2019 et selon une [recherche sur l'informatique du Cloud menée par IDG en 2018](#)¹, le reste devrait suivre d'ici à 2021. Mais le passage au cloud n'est pas une stratégie du tout ou rien. En fait, la plupart des entreprises, (et en particulier celles qui font déjà appel à des applications sur site existantes), optent pour une stratégie cloud hybride en déployant des applications, des données et une infrastructure sur une combinaison de ressources sur site et dans le cloud.

Le cloud hybride offre de nombreux avantages en termes de déploiement rapide et de réduction de coûts d'infrastructure, mais il s'accompagne de nouveaux défis en matière de gestion des données. Au fur et à mesure que les environnements dans le cloud se multiplient, de nouveaux silos de données apparaissent, et certains d'entre eux vont au-delà de l'informatique. La protection et l'administration des données présentes sur plusieurs clouds possédant chacun leur propre architecture sont des exercices compliqués. Et être dépendant d'un seul fournisseur cloud peut rendre la migration ou l'exportation des données difficiles.

En conséquence, votre stratégie (ou l'absence de stratégie) en matière de cloud peut s'avérer vraiment problématique lorsqu'il s'agit d'accéder à vos données, de les gérer, de les protéger et d'en tirer rapidement des informations. [Et comme 95 % des données enregistrées par des cadres dirigeants sont une partie intégrante de leur stratégie commerciale](#)², le temps est venu de vous assurer que c'est votre stratégie de données qui régit votre stratégie de cloud hybride, et non l'inverse.

Pourquoi adopter une stratégie de données d'entreprise ?

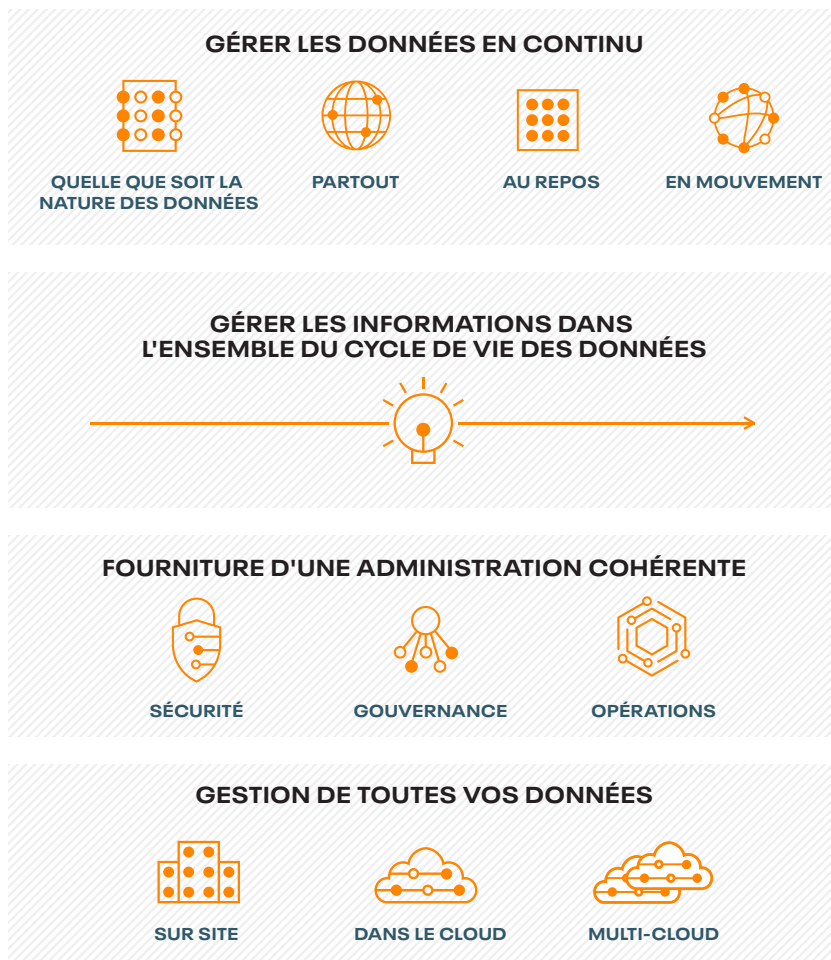
Piloter une organisation informatique pendant la transformation de votre organisation en architecture hybride bien gérée, exige de faire appel à de nouveaux modes de réflexion sur votre stratégie de données. En fait, il vous faut une stratégie de données plutôt qu'une stratégie cloud. Vos données sont un atout stratégique et elles doivent être traitées en tant que telles. Le cloud offre une importante opportunité pour un modèle évolutif et fiable, mais c'est une stratégie de données bien planifiée qui vous permet de maîtriser les coûts et de réduire vos risques grâce à la mise en œuvre de systèmes de sécurité et de gouvernance cohérents dans les actifs de données de votre entreprise.

Votre stratégie de données vous permet de limiter les risques en se concentrant sur le stockage, la gestion et la protection des données. Elle fournit sécurité et gouvernance, éléments indispensables pour limiter la fraude, éviter le vol et garantir la conformité sans lesquels l'entreprise pourrait s'exposer à des sanctions onéreuses, ou à une interdiction d'exercer. Elle garantit également l'intégrité et l'exactitude des données qui circulent dans votre entreprise, ce qui permet de les faire connaître, de les rendre visibles, disponibles, fiables et conformes. Votre stratégie de données doit également soutenir les objectifs de votre entreprise tels que l'augmentation des recettes, l'amélioration de la satisfaction client et la génération de profits. La découverte et la fourniture de données permettent à vos unités d'exploitation d'obtenir rapidement des informations, d'accélérer l'amélioration des produits ou l'expérience client et de comprendre comment réaliser des économies de coûts.

Importance d'une architecture de données moderne

Votre stratégie de données d'entreprise doit tendre avant tout vers l'établissement d'une architecture hybride offrant des services et des fonctionnalités de données cohérents, vous permettant de partager des données, des métadonnées et des charges de travail entre des environnements structurels, privés et publics.

Architecture de données moderne



Une architecture de données moderne vous permet de :

Utiliser le cloud comme une infrastructure.

Les fournisseurs cloud proposent une infrastructure essentielle afin de répondre à vos besoins dans le cloud, mais ne fournissent pas les capacités fondamentales nécessaires pour établir une architecture hybride au niveau de l'entreprise.

Donner la priorité à l'open source.

L'établissement d'une architecture ouverte sur les environnements locaux et du cloud public garantit la portabilité des applications sans reprise des programmes.

Trouver l'équilibre entre les besoins de l'entreprise et ceux de l'informatique

Les intervenants des unités d'exploitation souhaitent accélérer le processus d'information, et vous devez donc proposer un accès aux données en libre-service sur demande tout en

permettant aux acteurs de l'informatique de réduire les risques et de garantir la conformité aux normes de l'entreprise en termes de sécurité, de gouvernance de données et de fiabilité opérationnelle.

Assurer la sécurité et la conformité des données sensibles.

Définissez et mettez en œuvre un ensemble de stratégies de sécurité et de gouvernance dans des environnements de cloud hybrides, et notamment les contrôles d'accès ciblés, la traçabilité des données et les journaux d'audit.

Trouver l'équilibre entre les besoins de l'entreprise et ceux de l'informatique

Lorsqu'on transfère les charges de travail d'analyse des données vers le cloud, il est important de connaître les métriques de ressources clés qui permettent d'exécuter votre charge de travail, vos applications et vos jeux de données, ainsi que le modèle de facturation cloud et les frais de service associés. Cela fait partie des efforts que nous déployons en continu afin de planifier et rationaliser les charges de travail, les applications et les jeux de données qui conviennent au cloud et ceux qui restent sur site.

Cet article est conçu pour aider les Directeurs techniques (CTO), les Directeurs des systèmes d'information (CIO) et les Responsables en charge des données (CDO) à comprendre les aspects stratégiques et commerciaux de la gestion des données dans un environnement cloud hybride. Votre stratégie de données d'entreprise doit réussir à surmonter les défis du cloud hybride, en offrant en permanence une valeur métier, et permettre de gérer, sécuriser et gouverner les données présentes sur site dans les environnements multi-cloud et périphériques

1. Traiter le cloud comme une infrastructure, et non comme une architecture de données.

Les premiers adeptes du cloud ont commencé leur parcours avec un seul prestataire du secteur pour de bonnes raisons (la disponibilité géographique, les niveaux de service, entre autres) mais réalisent aujourd'hui que la stratégie de cloud d'entreprise et leur présence commerciale au niveau international imposent de faire appel à de nombreux fournisseurs cloud pour être disponible au niveau international. À la fin de la journée, le cloud vous fournit un accès à une infrastructure de calcul et de stockage, de manière comparable à une centrale électrique qui fournirait à ses clients un accès à l'électricité. Les entreprises devraient être en mesure de tirer profit des ressources de calcul et de stockage sans craindre d'être dépendantes de leur fournisseur ou d'une couche d'application propriétaire qui empêcherait leur progression dans le monde du cloud hybride.

Les fournisseurs cloud mettent à disposition une infrastructure indispensable à vos besoins sur le cloud, mais n'offrent pas les capacités fondamentales nécessaires à l'établissement d'une architecture de données au niveau de l'entreprise. Aujourd'hui, la plupart des organisations gèrent un écosystème d'application et de données très complexe, qui exige un éventail complet de fonctions d'entreprise telles que la sécurité et la gouvernance, les contrôles opérationnels, la portabilité des applications et l'assistance aux entreprises.

Vos prestataires du cloud ne sont tout simplement pas axés sur la mise en place d'une architecture de données hybride parce qu'ils ne disposent tout simplement pas de locaux physiques sur site (si ce n'est auprès d'autres prestataires du web), et qu'ils ont pour objectif de vous garder dans leur écosystème de calcul, de stockage, et d'applications. Donc, votre stratégie de données cloud hybride exige d'autres technologies pour répondre aux exigences de la gestion des données au niveau international tout en tirant le meilleur profit de votre investissement dans l'infrastructure.

2. Donner la priorité à l'open source

Lorsque vous entreprenez un parcours visant à définir une stratégie de données d'entreprise, pensez à vos actifs de données disponibles au sein de l'entreprise, à l'échelle mondiale, et englobez plusieurs environnements locaux, cloud et multi-cloud. Pour exploiter réellement le potentiel de vos données et les transformer en atout stratégique, vous aurez besoin d'une technologie des données open source capable de rassembler, de protéger et de gérer les actifs de données d'entreprise en silos, quel que soit l'endroit où elles se trouvent.

- de soulager vos inquiétudes en matière de « dépendance » d'un seul fournisseur,
- de profiter du rythme rapide de l'innovation généré par la communauté de logiciel open source,
- de tirer parti de partenariats de l'écosystème open source,
- et de garantir une réussite pour votre entreprise, sans dépendre d'une technologie propriétaire quelconque.

Tous les prestataires open source n'offrent pas une assistance intelligente capable d'analyser de manière proactive vos environnements de big data à la fois sur site et dans le cloud pour optimiser vos performances et réduire les risques. Les fournisseurs ne doivent pas seulement répondre dans le cadre du contrat de service indiqué, mais également offrir la sécurité et des correctifs en temps et en heure, et les traduire dans un code open source.

Votre prestataire d'assistance d'entreprise doit être un vrai partenaire, qui communique souvent dès le début, et s'arrange pour aligner les intérêts de votre entreprise avec la vision de la communauté de logiciel ouverte et la feuille de route du produit. L'innovation dans la communauté open source se fait par consensus, et c'est un vote qui décide d'inclure ou non des modifications de code. Recherchez un fournisseur qui a obtenu des votes favorables de la part de la communauté et qui n'est pas un simple fournisseur de packages de logiciels open source ou prestataire de service.

3. Équilibrer les besoins de l'entreprise et ceux de l'informatique

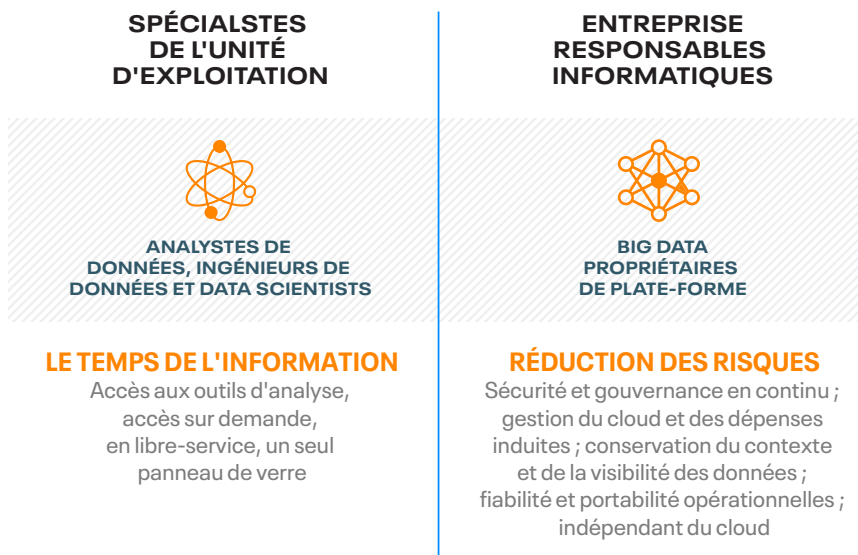
Lorsque vous entamez une réflexion sur une stratégie de données d'entreprise, il est indispensable de trouver un équilibre entre les besoins des spécialistes de votre secteur d'activité et les normes informatiques de l'entreprise. Votre stratégie de données doit permettre aux spécialistes des unités d'exploitation de raccourcir les délais d'obtention des informations grâce à un accès en libre-service sur demande et à une large gamme d'outils d'analyse. D'autre part, les intervenants de votre département informatique, au sein de l'entreprise, doivent garder le contrôle de la technologie de données d'entreprise, de façon à réduire les risques et à garantir la conformité aux normes informatiques d'entreprise en matière de sécurité, de gouvernance des données et de fiabilité opérationnelle.

Le développement d'une stratégie de données d'entreprise offrant aux intervenants de l'unité d'exploitation les outils dont ils ont besoin sans créer une « informatique fantôme » ou dupliquer des données et des silos d'analyse n'est pas une tâche facile. Les fuites de données et les infractions peuvent handicaper n'importe quelle entreprise, et les responsables informatique doivent avoir une vision à long terme d'une stratégie de données capable de simplifier l'intégration et les opérations, sans faire obstacle au progrès.

Dans la course au cloud et à l'agilité, la simplicité et l'efficacité qu'il génère, il est indispensable de connaître et de rationaliser vos actifs de données et les charges de

travail pour les adapter au cloud en prenant en compte la sécurité, la sensibilité, la propriété et les autres inquiétudes. Cet aspect est particulièrement important, car vous avez probablement investi des années dans la mise en œuvre, l'optimisation et l'exploitation des systèmes d'enregistrement dans les environnements sur site uniquement.

Équilibrage des exigences d'entreprise en matière de stratégie de données cloud



4. Assurer la sécurité et la conformité des données

La détermination des charges de travail devant rester sur site et de celles devant être transférées vers le cloud exige également une connaissance exhaustive de vos actifs de données. Par exemple, pour réduire le risque de violation de vos données, il se peut que vous ayez décidé de garder les données sensibles telles que les informations permettant de vous identifier (PII), les informations des cartes de paiement (PCI), les informations de santé protégées par HIPAA (PHI), et d'autres données de type réglementaire sur site, et de déplacer les données et charges de travail moins sensibles vers le cloud.

Cependant, la tendance actuelle est plutôt au transfert de données de plus en plus sensibles vers le cloud, ce qui nécessite la mise en place d'un modèle de sécurité unifié. Pour réduire les risques d'exposition de vos données et les accès non autorisés à ces dernières, vous devez avoir le contrôle de la sécurité et de la gouvernance sur site et sur le cloud, ce qui vous permet d'appliquer des stratégies de sécurité ultra ciblées, avec la provenance et la traçabilité des données de bout en bout, ainsi que des pistes d'audits permettant de savoir qui a accédé aux données. Après tout, en cas d'accès non autorisé, vous devez être immédiatement en mesure de repérer les actifs de données exposés, les informations susceptibles d'avoir fuité et d'identifier les coupables en analysant un journal d'audit des accès au niveau du système.

5. Gérer les coûts et optimiser les ressources

Connaître les frais de service du cloud peut être primordial si l'on considère que chaque service utilise un modèle tarifaire différent, composé de plusieurs éléments. Par exemple, le prix de certains services est calculé sur la base de la consommation des ressources de calcul nécessaires pour les machines virtuelles et les capacités de stockage. En fonction du support de stockage utilisé, le prix de l'entreposage peut être évalué en fonction de la quantité d'espace consommée ou du stockage fourni, quelle que soit la consommation réelle. D'autres services peuvent être tarifés en fonction des données numérisées sur la base d'un prix à la requête, du nombre de requêtes d'API, du nombre d'octets transférés sur un réseau, et de nombreux autres facteurs souvent difficiles à estimer à l'avance. Si votre environnement de données et d'analyse est inactif, ou si vos travaux, requêtes et pipelines ne sont pas correctement conçus et configurés, vous pouvez en réalité encourir des frais significatifs en raison d'une consommation inadéquate des ressources du cloud tout en les utilisant peu, ou sans tirer aucun profit commercial de cette consommation.

Vos pipelines de données peuvent intégrer plusieurs étapes, chacune nécessitant une technologie différente. Par exemple, un pipeline de traitement de données de bout en bout peut faire appel à des outils tels qu'Apache Kafka pour la diffusion en continu des données, Apache NiFi pour la gestion des flux de données, Apache Spark pour la data science et le machine learning (ML), Apache Hadoop pour le stockage des données à l'aide d'un système de fichier distribué Hadoop (HDFS), et Apache Hive pour l'exécution des requêtes SQL en vue de répondre aux questions de veille commerciale à partir de vos données.

La mise en œuvre de l'ensemble du pipeline avec des services natifs cloud peut s'avérer problématique, tant du point de vue opérationnel que des perspectives budgétaires. Certains prestataires cloud peuvent offrir des services natifs offrant des fonctionnalités équivalentes, mais il existe trois défis majeurs :

1. Les fournisseurs cloud regroupent le plus souvent des logiciels open source, et en général, ils n'ont pas les compétences open source nécessaires à la prise en charge de vos charges de travail de données en production, si l'on exclut la publication de correctifs et de mises à jour dans la communauté.
2. La responsabilité d'orchestrer et d'intégrer différents services natifs cloud dans un pipeline de traitement de données de bout en bout, sans aucune assistance au niveau de l'entreprise de la part du prestataire, vous incombe
3. Les modèles de facturation des prestataires du cloud de chaque service peuvent être complètement différents, ce qui rend le total des frais impossible à estimer à l'avance lorsqu'on transfère les charges de travail d'analyse des données vers le cloud, il est important de connaître les métriques de ressources clés permettant d'exécuter votre charge de travail, le modèle de facturation cloud et les frais de service associés. Une des principales stratégies de contrôle des frais consiste à garder les charges de travail à long terme sur site et à déplacer les charges de travail éphémères nécessitant une certaine élasticité des ressources sur le cloud. Disposer des contrôles opérationnels convenant à l'optimisation des ressources cloud pour vos charges de travail d'analyse peut vous faire réaliser des économies substantielles.

About Cloudera

Chez Cloudera, nous pensons que les données peuvent rendre ce qui est impossible aujourd'hui possible demain. Nous donnons aux personnes la possibilité de transformer des données complexes en informations claires et exploitables.

Cloudera offre un cloud de données d'entreprise pour n'importe quelle donnée, partout, from the Edge to AI. Alimentée par la perpétuelle innovation générée dans la communauté de l'open source, Cloudera contribue à faire progresser la transformation numérique pour les plus grandes entreprises du monde.

En savoir plus, sur cloudera.com

Connectez-vous avec Cloudera

À propos de Cloudera :
cloudera.com/more/about.html

Consultez notre blog VISION :
vision.cloudera.com
et notre blog Engineering :
blog.cloudera.com

Suivez-nous sur Twitter :
twitter.com/cloudera

Rendez-nous visite sur Facebook :
facebook.com/cloudera

Venez nous voir sur YouTube :
youtube.com/user/clouderahadoop

Rejoignez la communauté Cloudera :
community.cloudera.com

Ayez un aperçu des réussites de nos clients :
cloudera.com/more/customers.html

Étapes suivantes

Actuellement, les architectures de données doivent prendre en charge des environnements de cloud hybrides, mais pas après coup. C'est au contraire votre stratégie de données qui doit mener la stratégie cloud afin que les deux soient alignées et se supportent l'une et l'autre. La combinaison d'un cloud hybride avec diverses capacités d'analyse sur une plate-forme unique est également connue sous le nom d'Enterprise data cloud. Un Enterprise data cloud représente une nouvelle architecture de gestion des données qui inclut :

- L'analyse multifonctions prévoyant le traitement des données et l'analyse à la périphérie de l'entreposage de données, des opérations en temps réel, et de machine learning
- Une sécurité, une gouvernance, des métadonnées et contrôles permettant de maîtriser votre stratégie de données d'entreprise
- Des infrastructures ouvertes pour l'intégration à des outils de tiers
- Et la flexibilité nécessaire au déploiement des cas d'usage sur les clouds publics, privés et hybrides exploitant n'importe quel type de données où que se trouvent ces dernières.

Pour obtenir davantage d'informations et une infrastructure pour garantir les opérations, téléchargez "Lancer la transformation de votre entreprise grâce à Enterprise Data Cloud."³

Sources

- ¹ https://resources.idg.com/download/executive-summary/cloud-computing-2018?utm_campaign=Cloud%20Computing%20Survey%202018&utm_source=Forbes%20-%20Louis
- ² <https://www.edq.com/globalassets/white-papers/2018-global-data-management-benchmark-report.pdf>
- ³ <https://www.cloudera.com/content/dam/www/marketing/resources/whitepapers/power-your-business-transformation-with-edc.pdf.landing.html>