

Les Données pour stimuler l'innovation à l'ère numérique avec Cloudera sur Microsoft Azure

Transformation numérique et blocs d'innovation commerciale

La transformation numérique (DX) est devenue essentielle au succès. IDC estime que d'ici 2022, plus de 60% du PIB mondial sera numérisé, la croissance de chaque secteur industriel sera liée à de meilleures offres numériques et près de 7 mille milliards de dollars seront investis en informatique.

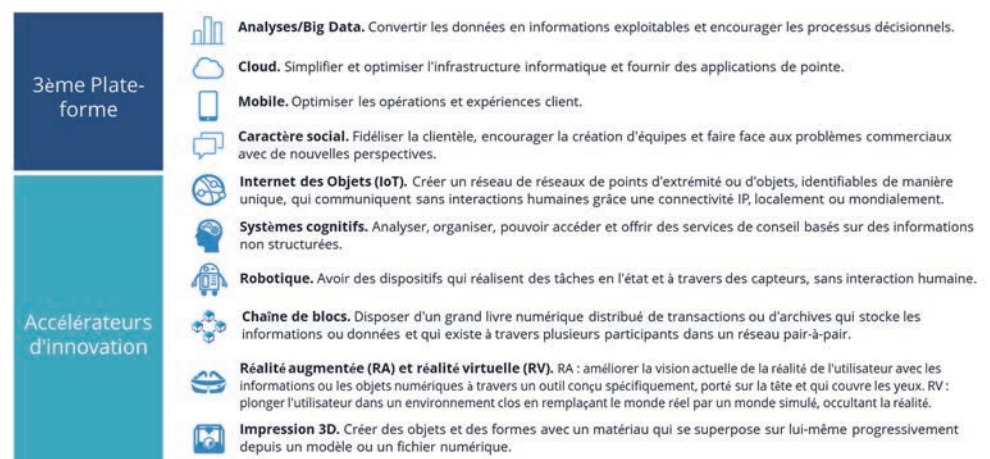
La transformation numérique est, de sa nature, une technologie et une stratégie d'affaires basée sur les données. Les entreprises veulent améliorer les expériences client et l'efficacité opérationnelle, identifier de nouvelles sources de revenus et construire une marque numérique de confiance, en convertissant leurs données en élément vital. Ainsi, la transformation des données est essentielle à la transformation numérique.

Les technologies fondamentales de cette transformation numérique sont le Cloud, le Big Data, la mobilité et le modèle social (collaboratif) ; IDC les nomme « les piliers de la 3ème plate-forme ».

Ces dernières années, ces piliers ont permis aux organisations de transformer leurs infrastructures et applications.

Mais nous entrons maintenant à la phase suivante de la transformation numérique où les organisations veulent associer à la 3ème plate-forme des « accélérateurs de l'innovation » comme l'IoT, les systèmes cognitifs (intelligence artificielle, apprentissage automatique, apprentissage profond), la robotique, les chaînes de blocs, la réalité augmentée et la réalité virtuelle, l'impression 3D et la sécurité de dernière génération, pour innover à échelle.

Schéma 1 Technologies révolutionnaires

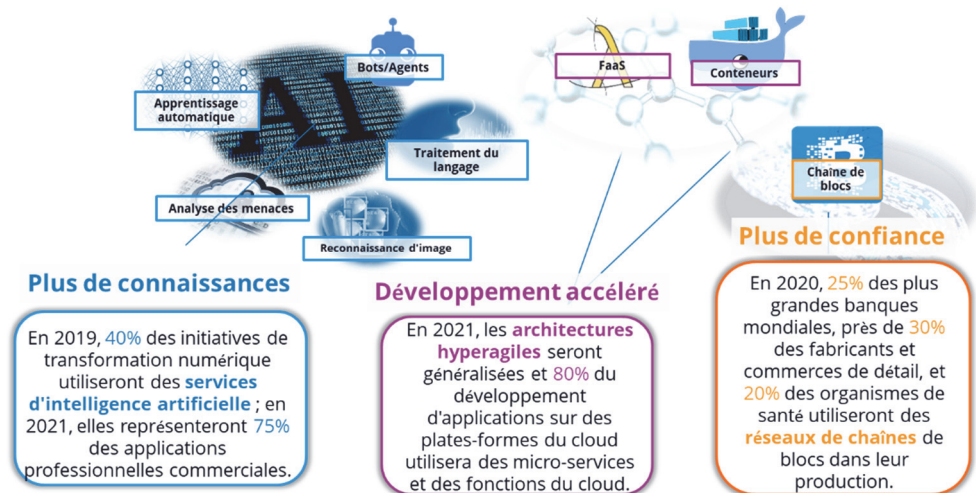


Source : IDC, 2019

Une plate-forme d'analyse des données qui fonctionne sur plusieurs plates-formes du cloud possède les fonctionnalités de base de gestion globale des données, de sécurité et d'analyses. Elle compte alors sur des fonctionnalités modernes d'entreposage des données qui peuvent servir de base solide pour la stratégie de données d'une entreprise et ainsi l'aider à mettre en œuvre des initiatives d'IoT, d'intelligence artificielle, d'apprentissage automatique, et de science des données, et ce de manière durable.

Cette deuxième phase de la transformation numérique est une ère d'« innovation accélérée » dans laquelle l'infrastructure hyperéchelle (hyperscale), l'intelligence artificielle, l'IoT, les conteneurs, les FaaS et chaînes de blocs seront les technologies déterminantes pour accélérer l'innovation et le développement. Cela signifie qu'il faut absolument adapter la stratégie informatique et des données à une infrastructure de cloud distribué, à des piles logicielles de cloud public, à un développement d'applications agiles et natives du cloud, et à la science des données.

Schéma 2
Technologies accélératrices d'innovation



Source : IDC, 2019

Les indispensables pour tirer profit de toutes vos données

Les données, d'abord simple outil professionnel, évoluent pour devenir indispensables à notre vie . D'ici 2025, près de 20% des données mondiales seront essentielles à notre quotidien et près de 10% de ces données seront absolument cruciales. À l'ère numérique, on dit que les entreprises sont, en premier lieu, des entreprises de logiciel et technologie. Cela dit, on peut affirmer qu'elle sont aussi des entreprises de l'information. IDC a identifié quelques tendances majeures qui renforcent le rôle des données :

- IoT et systèmes intégrés. Les appareils connectés vont générer d'énormes quantités de données qui permettront d'affiner et d'améliorer les systèmes et processus d'une manière jusque-là inimaginable. De nouveaux rôles spécifiques liés aux données vont apparaître. Le Big Data et les métadonnées (données sur les données) deviendront alors très importants.
- Hyperdisponibilité des données et données mobiles en temps réel. De plus en plus, les données devront être disponibles instantanément, partout et à toute heure. Les entreprises de tous les secteurs à travers le monde vivent une transformation numérique (DX) motivée par ces besoins. D'ici 2025, plus du quart des données créées mondialement sera généré en temps réel et les données de l'IoT en temps réel représenteront 95% de cette portion.

- Les systèmes d'intelligence artificielle qui changent la donne. Le flot de données permet à un nouvel ensemble de technologies (qui réunit l'apprentissage automatique, le traitement de langage naturel et l'apprentissage approfondi - connus sous le nom de systèmes d'intelligence artificielle) de transformer l'analyse des données en outil proactif de prise de décision et de mesure stratégiques. IDC estime que, d'ici à 2025, 50 fois plus de données seront analysées pour un total de 5,2 ZB, et les données analysées par un système d'intelligence artificielle seront 100 fois plus nombreuses qu'aujourd'hui, pour un total de 1,4 ZB.
- Sécurité avant tout. Les données provenant de nouvelles sources entraînent plus de vulnérabilité des informations personnelles et sensibles. Il existe un fossé majeur entre la quantité de données produites aujourd'hui qui exigent de la sécurité et la quantité de données qui sont vraiment protégées ; un fossé qui ne cessera de s'élargir dans notre monde basé sur les données.

Rien qu'en Europe occidentale, IDC prévoit que le marché des logiciels du Big Data et des analyses (BDA) va croître à un taux de croissance composé annuel (TCCA) de 9,2% d'ici 2022. Dans ce même marché, on prévoit que le segment des outils de BI et d'analyses et celui des plates-formes de gestion et d'intégration des données analytiques, enregistrent un TCCA de 10%.

Le marché du BDA observe une forte croissance grâce aux initiatives orientées données et aux trois changements majeurs qui se produisent sur le marché :

- La croissance accélérée du logiciel de gestion des données, avec des bases de données non relationnelles, ainsi que la tendance vers un IoT auquel on a associé des fonctionnalités d'analyse et d'apprentissage automatique pour faciliter le traitement et la gestion des données de l'IoT afin de prendre des décisions éclairées. Le marché EMEA d'outils de BDA et IoT devrait enregistrer un TCCA de 30,7% d'ici 2021 car il suit le courant dominant.
- La migration vers le cloud public. Le fait de rapprocher les données du lieu où elles ont été générées pour réduire les frais généraux oblige les organisations à considérer des déploiements dans le cloud car c'est là que se trouvent les applications modernes.
- La demande croissante d'automatisation se reflète dans le besoin majeur en systèmes cognitifs ou d'intelligence artificielle. Les entreprises voient déjà l'automatisation, les analyses et l'apprentissage automatique comme des éléments essentiels grâce à leur capacité à automatiser, standardiser et générer plus d'information cruciale dans des tâches répétitives et routinières, ce qui peut améliorer la productivité et optimiser l'utilisation des ressources disponibles pour des tâches stratégiques et plus pertinentes.

Monétariser les données est devenu essentiel pour créer un avantage concurrentiel. L'analyse en temps réel est une des charges de travail les plus rapides, selon la recherche européenne menée par IDC en 2018 sur des utilisateurs finaux

multicloud. Cependant, pour les organisations européennes, cela doit se produire dans un contexte de respect du RGPD et des obligations en matière de confidentialité des données, ce qui n'est pas si simple.

Le sondage mené par IDC en 2017 sur la valeur des données, plus précisément sur la capacité des organisations à tirer profit des données, révèle que seule une minorité en profite pleinement (11%) alors que près des trois quarts (74%) « résistent » aux données, les utilisent peu ou comme elles peuvent, laissant ainsi un large fossé entre les organisations qui utilisent les données comme avantage concurrentiel et celles qui ne le font pas.

Défis liés à la clientèle

IDC estime que les données mondiales atteindront 163 ZB en 2025. La grande quantité et variété de sources de données, les exigences légales plus strictes, la stagnation des budgets et la fragmentation des données à travers le cloud multiple sont des défis à relever pour transformer les données en informations exploitables. Dans leurs conversations avec IDC, les grandes entreprises admettent que l'évolutivité verticale est un échec et qu'elles regroupent les dispositifs dans des clusters pour traiter les données avec des systèmes de traitement distribué.

Il est nécessaire d'avoir un registre de données et une plate-forme de diffusion en continu de dernière génération, et de rompre avec les multiples entrepôts de données traditionnels répartis en silos, de manière à ce que les requêtes obtiennent des réponses instantanées et plus fréquentes et que la gestion soit simplifiée. Au-delà de cet aspect, l'apprentissage automatique joue un rôle clé en fournissant de meilleures informations liées à l'excellence opérationnelle, aux processus opérationnels, au marketing et aux applications de sécurité.

En dehors de l'aspect technologique, on peut ajouter d'autres défis tels que le déficit de compétences, la quantité limitée de scientifiques des données et les budgets limités alloués à l'informatique pour l'innovation.

Les premières entreprises à avoir adopté les analyses du Big Data doivent aussi faire face aux stratégies d'analyses traditionnelles de l'ère antérieure au multicloud.

1. Silos de données : données stockées dans des infrastructures du multicloud
2. Qualité des données : données manquantes et non vérifiées
3. Informatique de l'ombre (Shadow IT) : les scientifiques des données et les professionnels du Big Data choisissent leurs propres outils, créant alors de nouveaux facteurs de vulnérabilité
4. Hétérogénéité des outils d'analyse : plusieurs outils à fonction unique utiles pour une seule infrastructure ou des tâches nécessitant des efforts d'intégration
5. Transformation informatique : les utilisateurs informatiques à l'ère du cloud qui n'ont plus les mêmes attentes et l'effort réalisé pour que la transition du cloud soit rapide et avec un risque limité

« Coudera sur Azure est la solution idéale pour nous, même pour nos nouvelles initiatives comme l'IoT », affirme Sven Loeffler, responsable des initiatives orientées données chez Deutsche Telekom.

Dans ce contexte, IDC considère que le partenariat entre Microsoft Azure et Cloudera Hortonworks est déterminant pour créer un ensemble de technologies génériques et de modèles commerciaux afin d'aider les organisations européennes dans la transition vers le cloud public et d'améliorer la compétitivité de leurs données. Grâce à cette alliance, les entreprises pourront tirer profit des plates-formes analytiques dans lesquelles elles ont déjà investi, les déplacer vers une plate-forme de cloud public et les utiliser comme une opportunité de moderniser leurs stratégies d'analyses et devenir alors des coordinateurs de données (data stewards) dans l'ère numérique.

Stratégies de capitalisation des données : les clés de la réussite

Le Cloud est un tremplin de l'innovation et de la mise en marché rapide : un objectif clé pour les entreprises. Le Big Data a beaucoup changé avec l'évolution du cloud et la valeur croissante des données : jusqu'à aujourd'hui, seuls les services informatiques étaient rattachés aux données, décidaient de la technologie, de l'infrastructure et des systèmes de l'entreprise.

Cependant, grâce au cloud et à un accès simplifié à l'infrastructure, il existe aujourd'hui beaucoup plus de personnes qui influencent les décisions informatiques : les propriétaires d'applications, les utilisateurs commerciaux, les scientifiques des données et les développeurs.

Le plus important est, d'une part, de développer une culture au sein de l'entreprise pour traiter les données comme un atout précieux et, d'autre part, d'investir dans des compétences et dans la science des données pour tirer profit des informations pertinentes avec des mesures adéquates de sécurité, confidentialité et protection des données.

Donner plus de pouvoir aux scientifiques des données et leur permettre de choisir leurs outils librement représente une opportunité majeure de succès dans la capitalisation et l'analyse des données. Ces nouveaux personnages qui créent des informations de valeur à partir des données d'entreprise les orientent vers les outils et services dans lesquels elles devraient investir. Lors d'échanges avec IDC, les scientifiques des données affirment qu'ils préfèrent les outils qui apportent plus d'autonomie et de flexibilité de déploiement offrant ainsi de grandes capacités allant de l'informatique en périphérie (edge computing) à l'informatique en nuage (cloud computing), avec une intelligence artificielle intégrée pour préparer l'avenir.

Parallèlement, cela crée la base technique pour une infrastructure du cloud agile, évolutive, sécurisée, à code ouvert et basée sur les API, qui s'intègre avec les plates-formes d'analyses que choisissent les clients.

Plate-forme numérique avec un noyau intelligent

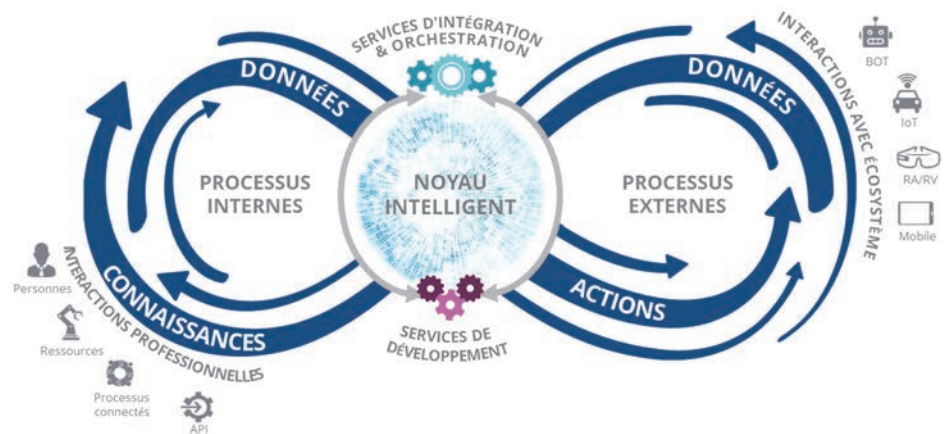
Les entreprises doivent penser différemment leurs architectures pour réussir à innover rapidement à échelle. IDC prévoit que d'ici la fin de l'année prochaine, 60% de toutes les entreprises mettront en place une nouvelle base informatique.

Le schéma 3 est une représentation conceptuelle de ce qu'est une plate-forme numérique et de comment elle est organisée. Cette plate-forme est définie par un logiciel, elle est orientée API et données, automatisée et hyperéchelle. Elle facilite

l'utilisation des flux de données externes et les associe aux données et informations internes, à l'aide de technologies comme l'intelligence artificielle et l'apprentissage automatique, afin de répondre à ces flux de façon numérique. Sécurité, automatisation, intégration et orchestration s'ajoutent à ce noyau.

Schéma 3

Plate-forme numérique avec un noyau intelligent



Source : IDC, 2019

La plate-forme de transformation numérique possède un « noyau intelligent », où les données sont gérées et utilisées et où se trouvent les algorithmes, le code de programmation et les modèles. Le noyau intelligent contient des données qui peuvent se convertir en valeur ajoutée pour l'entreprise, au-delà des systèmes et des processus individuels. Puisque l'innovation se produit le plus souvent en périphérie (edge), il est essentiel de pouvoir accéder aux données de façon simple, depuis n'importe où dans l'écosystème. Face à l'énorme flux, l'analyse des données est en train de devenir une méthode proactive pour la prise de décisions et de mesures stratégiques, aux nouvelles technologies disponibles comme l'apprentissage automatique, le traitement de langage naturel et l'apprentissage approfondi, plus connus sous le nom de systèmes d'intelligence artificielle.

Une autre étape importante consiste à faire évoluer les stratégies d'entreposage des données et de Big Data pour relever les défis de fragmentation des données, d'absence d'une vision globale et d'augmentation des coûts d'analyse. Les entreprises doivent évaluer les outils modernes d'entreposage des données qui aident à les analyser depuis de nouvelles sources et à analyser les données non structurées pour offrir une vision complète des données pertinentes à transformer en informations exploitables.

Les organisations doivent aussi développer des stratégies à long terme pour être orientées données et accéder aux fonctionnalités avancées d'analyse, aux analyses en temps réel et au spectre complet allant de l'IoT à intelligence artificielle ; cela pour décider de l'évolution de la science des données au sein de l'entreprise.

Les bases que les entreprises choisissent aujourd'hui pour la transformation de leurs données détermineront leur succès de demain. Selon IDC, l'agilité, l'indépendance de la plate-forme, l'ouverture, la richesse de l'écosystème, les

intégrations approfondies, l'innovation technologique, la crédibilité du vendeur, et même les histoires des clients sont des considérations importantes.

Pour capitaliser leurs données, les directeurs des systèmes d'information et les responsables informatiques doivent :

- Développer et mettre en place une base technique industrielle de capitalisation des données
- Clarifier le rôle et les attentes des organisations informatiques concernant la gestion et la monétisation des données
- Impliquer les parties prenantes tels que les scientifiques des données, les ingénieurs, les analystes du Big Data, les experts en sécurité et la communauté de développeurs pour construire la stratégie de données

Comprendre le monde en rapide évolution du Big Data et des analyses

Quand on parle d'analyse des données, on emploie plusieurs termes différents mais similaires aussi, qu'il vaut donc mieux clarifier pour éviter les confusions. IDC estime qu'en fin de compte, les entreprises doivent comprendre les problèmes qu'elles veulent résoudre ou identifier leurs utilisations pour choisir la bonne initiative de capitalisation des données et atteindre leur objectif, plutôt que de suivre tête baissée le mouvement du Big Data.

1. **Analyse.** Pratique générale qui consiste à soutirer des renseignements utiles des données à l'aide d'outils, de technologies et de méthodes. Elles peut être descriptive, prédictive ou prescriptive.
2. **Science des données.** Similaire à l'analyse, la science des données se trouve à l'intersection entre la statistique et l'informatique. Elle se réfère aux objets, techniques et méthodes d'analyse utilisées par les scientifiques des données, des professionnels avec une compréhension approfondie des algorithmes et technologies d'analyse.
3. **Apprentissage automatique.** C'est un ensemble d'outils et de techniques utilisés pour entraîner les systèmes informatiques à générer des prédictions sans qu'on leur fournisse des instructions explicites. L'apprentissage automatique peut s'appliquer à une grande variété de techniques allant de l'analyse de régression et du regroupement aux réseaux neuronaux.
4. **Exploration des données (Data mining).** C'est l'extraction de schémas, de relations et de connaissances depuis une grande quantité de données. Elle se trouve entre la statistique, l'apprentissage automatique et les systèmes de base de données.
5. **Intelligence artificielle.** Terme autrefois utilisé pour décrire l'utilisation de réseaux neuronaux mais aujourd'hui utilisé comme terme générique pour se référer à tout ce qui rend les machines intelligentes et automatisées.

Voyons maintenant comment l'offre de Microsoft et Cloudera s'ouvre à de nouvelles opportunités pour que les clients soient orientés données.

Offre de valeur de Microsoft Azure et Cloudera

Tous les regards se sont portés sur Microsoft, cinq ans plus tôt, quand son PDG Satya Nadella a promis une stratégie qui donnerait la plus grande priorité au cloud, à la mobilité et au code en open source. La société a beaucoup avancé depuis. Sur le front de l'open source, Microsoft a ouvert son portefeuille de brevets en rejoignant le réseau Open Invention Network (OIN), elle a forgé des alliances solides avec les distributeurs de technologie en open source et a développé des intégrations approfondies entre Azure et les systèmes opératifs en open source, les bases de données, les plates-formes analytiques et les charges de travail.

En lien avec l'innovation rapide dans le marché du cloud, Microsoft est en avant de la course sur plusieurs fronts comme les produits et fonctionnalités de sécurité, les outils pour développeurs, les fonctionnalités Devops, l'intelligence artificielle et l'IoT, ainsi que sur une grande variété de matériel et d'améliorations de services dans le cloud de Microsoft Azure. Elle continue aussi d'intégrer son large portefeuille de logiciels à ses services de cloud pour offrir des solutions de cloud hybride aux clients. Lors de sa récente conférence Ignite, Microsoft a voulu apaiser les inquiétudes liées au cloud public : la société a annoncé une amélioration et une plus grande gamme d'outils, a facilité le processus d'adoption d'une stratégie hybride et a augmenté la confiance portée au cloud public.

Microsoft a aussi investi dans des intégrations approfondies avec les meilleures plates-formes complémentaires du marché, comme Cloudera. Cette alliance donne aux clients la possibilité de profiter d'outils familiers d'analyse de données, de progresser dans leurs initiatives orientées données d'intelligence artificielle et d'apprentissage automatique, et de les étendre au cloud public. La combinaison des fonctionnalités d'Azure - Azure Data Factory, les services de données, les fonctionnalités de sécurité, les outils de gestion de données et l'assistance technique - avec les options modernes de Cloudera - entreposage des données, analyses basées sur l'intelligence artificielle et la gestion des données de bout-en-bout - peut aider les entreprises à progresser dans leurs objectifs de données.

Un des points forts du partenariat est la simplicité : les clients n'ont qu'un seul point de contact pour leur plate-forme de données Cloudera sur Azure et cela leur permet d'accélérer leur vitesse de travail. La simplicité se retrouve aussi dans le fait de réunir des données très fragmentées dans une unique plate-forme pour des analyses à fort impact.

D'un point de vue commercial, l'échelle et l'économie de l'infrastructure du cloud permet aux scientifiques des données d'expérimenter plus rapidement : ils créent des modèles avec de petits sous-ensembles sur la plate-forme de Cloudera et les adaptent ensuite aux services de données et aux fonctionnalités de sécurité d'Azure.

Plate-forme de données Cloudera (Enterprise Data Hub)

La solution réunit dans une plate-forme unique plusieurs disciplines d'analyses : les entrepôts de données, la science des données, l'ingénierie des données, la recherche, les analyses en continu et en temps réel. Cela améliore l'expérience des

données partagées et simplifie les changements de politique, la sécurité, la gouvernance de l'information et la gestion du cycle de vie des données.

Ce qui différencie la plate-forme de Cloudera, c'est sa multifonctionnalité : elle permet d'effectuer plusieurs fonctions au sein du même registre de données et de les réutiliser à travers d'autres plates-formes sans devoir dépendre d'un seul vendeur. IDC constate aussi que de multiples facteurs comme la conformité, la gouvernance, le coût, la sécurité et la proximité avec les charges de travail déterminent l'emplacement des données (ou optimisation de la localisation), ce qui oblige les entreprises à évaluer des technologies qui fonctionnent à travers ces infrastructures hétérogènes de multicloud et ces environnements de cloud hybride.

Cloudera a entrepris une transition avec pour élément central l'intelligence des données, allant au-delà d'Hadoop et devenant ainsi fournisseur de plates-formes de données. Elle s'est convertie en entreprise leader du secteur de l'intelligence artificielle, de la science des données et de l'apprentissage automatique pour les entreprises. Par ailleurs, elle révolutionne l'entrepôt de données pour les charges de travail analytique avec des fonctionnalités d'entreposage de données qui incluent des analyses de données structurées, semi-structurées et non structurées, et elle améliore en permanence la plate-forme pour simplifier les environnements complexes de données. Selon IDC, une plate-forme d'analyses de données qui fonctionne à travers plusieurs plates-formes de cloud mais aussi sur une infrastructure locale peut offrir aux clients une expérience de gestion et de visibilité unifiée des données. La plate-forme Cloudera possède les fonctionnalités clés de gestion des données de bout-en-bout, de sécurité, d'analyse,

et d'entreposage des données, qui peuvent servir de base dans la création d'une stratégie de données d'entreprise et de son développement durable avec l'IoT, l'intelligence artificielle, l'apprentissage automatique et la science des données, dans le cloud public mais aussi sur une infrastructure locale. Cette expérience unifiée est essentielle car elle donne aux entreprises la flexibilité de déplacer des charges de travail localement ou dans le cloud en fonction de leurs besoins et sans compromettre ou influencer l'expérience client.

Grâce à son investissement, ses efforts en ingénierie, ses fusions et partenariats, Cloudera est en passe de devenir la plate-forme de données de référence des clients. En premier lieu, sa fusion avec Hortonworks lui apporte des technologies complémentaires qui offrent des fonctionnalités allant de l'IoT à l'intelligence artificielle, simplifiant ainsi le parcours de transformation des données pour ses clients. En second lieu, la fusion augmente la proposition de valeur de Cloudera à ses clients : une plate-forme de gestion et d'analyses des données unifiée qui comprend une offre allant de l'informatique en périphérie à l'intelligence artificielle. Du point de vue de l'innovation, elle a ajouté des fonctionnalités modernes d'entreposage et de gestion des données.

Son partenariat avec Azure démontre son engagement d'intégrer ses technologies aux plates-formes hyperéchelle à forte croissance et montre comment elle tire profit des tendances de migration vers le cloud.

Partenariat à long terme

Microsoft Azure et Cloudera se sont unies en 2014 et ont favorisé leur intégration ces quatre ou cinq dernières années. Il existe des équipes de gestion de partenariat dédiées et les deux entreprises se considèrent comme associées de même niveau, ce qui indique un engagement profond pour que les technologies fonctionnent avec fluidité, améliorant ainsi l'expérience client.

Par exemple, Cloudera est disponible sur la Marketplace d'Azure, il suffit juste de l'activer en un clic.

Les éléments clés de cette collaboration sont les suivants :

- Azure IaaS. Les clients des deux entreprises peuvent profiter des avantages de base du cloud public comme les grandes fonctionnalités de calcul, le stockage d'Azure et son réseau Express Route pour avoir un accès fermé à l'environnement hyperéchelle et créer des instances sur la plate-forme de Cloudera pour les analyses.
- Expérience de science des données à travers un cloud hybride.
- Innovation de chaque vendeur avec intégration plus rapide. Des offres telles que l'apprentissage automatique de Cloudera sont prévues sur Azure, ainsi que l'intégration avec la plate-forme de conteneurs d'applications d'Azure Kubernetes Service. IDC voit une synergie inhérente à la stratégie de conteneurisation d'Azure et la vision d'applications natives du cloud de Cloudera.
- Vision complémentaire. Les deux entreprises ont des stratégies open source similaires : d'une part, Microsoft a fermement démontré son engagement envers le code ouvert ces deux dernières années, d'autre part, Cloudera est exécuté dans des environnements professionnels de Linux et la fusion avec Hortonworks renforce davantage sa stratégie open source. L'objectif principal de l'alliance Microsoft-Cloudera est de faciliter le parcours du client dans tout ce processus, et les entreprises financent conjointement les démonstrations de faisabilité et les phases liées à la migration. Les technologies open source comme Hadoop, les systèmes Linux et Cloudera sont profondément intégrées dans Azure. Microsoft crée un impact fort au sein de la communauté open source et apporte plus de fonctionnalités Azure aux plates-formes à code ouvert.

En 2018, Cloudera a été nommée, par Microsoft, partenaire de l'année pour son offre de données en open source et d'intelligence artificielle. IDC considère qu'il existe une convergence naturelle en terme de clients entre ces deux entreprises parce qu'elles se focalisent principalement sur les grosses sociétés.

Étude de cas

Cet exemple montre comment un client européen a utilisé Cloudera et, par la suite, la plate-forme, sur Microsoft Azure pour obtenir des bénéfices en interne grâce à des perspectives analytiques, offrir des services à ces clients et augmenter ses recettes.

Deutsche Telekom

La société de Télécoms allemande utilise le Big Data et les analyses pour les principales raisons suivantes : réduire les coûts des processus internes et de l'efficacité des systèmes, offrir de nouvelles opportunités aux employés et clients, et améliorer les cas d'étude comme la détection des fraudes.

En terme de réseau, la fraude est un défi majeur à relever pour les acteurs des télécommunications. Pour l'éviter et identifier des événements suspects, d'énormes quantités de données de réseau doivent être collectées et analysées.

La fragmentation des données de réseau devenait un gros obstacle aux analyses complètes, ce qui limitait la visibilité et rendait impossible l'apprentissage automatique à échelle. Elle a choisi Cloudera pour moderniser sa plate-forme de données car elle a la capacité d'analyser de larges ensembles de données, la possibilité d'exécuter l'apprentissage automatique et les analyses en temps réel, et d'offrir une expérience des données partagée.

L'entreprise a utilisé Cloudera Altus Data Warehouse, Cloudera Data Science Workbench et SDX (Expérience des données partagées) pour établir son Telekom Data Intelligence Hub : une plate-forme fiable, totalement gouvernée et un écosystème qui permet aux utilisateurs d'échanger et d'analyser des données, et de développer des applications multifonctions à partir des données, en toute sécurité.

La plate-forme moderne d'intelligence des données apporte des résultats et Deutsche Telekom a réussi à surmonter le défi des fraudes de réseau. La qualité du réseau et la détection des activités frauduleuses en temps réel ont été améliorées grâce à la grande échelle, au traitement rapide d'ensembles de données complets et à l'outil de requête interactif de Cloudera, ce qui a permis d'éviter 10%–20% de pertes en recettes.

La société de Télécoms a aussi obtenu une vision d'entreprise des données, allant de données de réseau aux données CRM, et les analyses lui ont permis de mieux répondre aux attentes de ses clients et d'améliorer sa qualité de service, de manière instantanée. L'initiative orientée données a aussi augmenté l'efficacité opérationnelle de 50%.

Tout a commencé en 2013-2014, quand DT a choisi Cloudera pour sa présence en Europe, son service et ses fonctionnalités modernes d'entreposage de données dans la plate-forme. Elle a démarré avec une distribution d'Hadoop sur une infrastructure privée dédiée sans systèmes opératif (bare metal). Cependant, quand le cloud public a commencé à gagner en popularité, DT a évalué le marché hyperéchelle et a découvert un allié naturel chez Microsoft Azure en 2016. Deux facteurs ont favorisé cette synergie idéale : un accord d'administration des données de T-Systems (appartenant à Deutsche Telekom) avec Microsoft Azure pour le marché européen, conforme aux normes sur la souveraineté et la confidentialité des données, et l'intégration de sa plate-forme de données existante, Cloudera sur Azure, et les bénéfices combinés de l'économie du cloud, des analyses modernes et de l'entreposage de données.

"Maintenant qu'elle est encore plus forte grâce à sa fusion avec Hortonworks, Cloudera sur Azure est notre solution préférée, même pour les nouvelles initiatives comme l'IoT", a déclaré Sven Loeffler, responsable des initiatives orientées données chez Deutsche Telekom.

La société cherche maintenant à agrandir son environnement Azure avec un cluster dédié aux analyses, plutôt qu'avec des clusters permanents uniquement. En se basant sur son expertise et son expérience, Deutsche Telekom utilise Cloudera sur Azure pour offrir à ses clients des services de gestion d'analyses de données et les accompagne dans leurs stratégies de transformation des données. « La transformation des données devient une priorité clé même pour les PME des marchés allemands et européens » ajoute Loeffler. IDC affirme que de plus en plus d'entreprises veulent être orientées données mais les petites et moyennes entreprises manquent des capacités internes pour adopter des solutions modernes d'entreposage des données et tirer profit des fonctionnalités d'intelligence artificielle et d'apprentissage automatique, de manière efficace, ouvrant ainsi un grand marché de services de gestion d'analyses de données pour des vendeurs comme DT. Cela augmente en retour les opportunités de marché de Cloudera et Azure. Lors d'échanges avec IDC, Cloudera et Azure ont mis l'accent sur l'importance du canal et de l'écosystème pour offrir des services de gestion d'analyse des données.

« La collaboration tripartite entre DT, Microsoft Azure et Cloudera est très forte et elle entreprend actuellement une phase bêta pour tous les nouveaux services », explique Loeffler. Le principal avantage de Cloudera sur Azure est l'incroyable amélioration de la vitesse de mise sur marché pour Deutsche Telekom.

« La vitesse de mise sur marché est de grande importance pour nous. Avant, il fallait compter un à trois mois pour fournir Cloudera localement sur notre infrastructure dédiée. Aujourd'hui, Cloudera sur Azure est disponible à la demande », déclare Sven Loeffler, Deutsche Telekom

La version sur demande dans le cloud a joué un rôle spécifique pour la société de Télécoms : elle a augmenté une de ses charges de travail pour interchanger des données au sein de l'entreprise et a dû rendre les données permanentes pour les utiliser dans la construction de nouveaux modèles. Cependant, la mise en place systématique de données permanentes revenait cher à cause du manque de prédictibilité de la charge de travail.

Avec la solution sur demande de Cloudera sur Azure, l'entreprise peut construire des modèles en fonction de la flexibilité des scientifiques des données et des développeurs. « Nous réduisons largement les coûts de cette manière », déclare Loeffler.

Les services de gestion d'analyses des données de DT sont populaires chez les entreprises automobiles et les PME.

Utilisation croissante et résultats commerciaux de l'analyse et la science des données

De plus en plus de charges de travail se prêtent aux analyses d'intelligence artificielle et à la science des données. Les applications de prédiction touchant à la logistique, la détection des anomalies et la détection des tendances sont en augmentation. On peut aussi compter les applications de santé et de médecine prédictive, la maintenance prédictive et la détection des fraudes dans les secteurs automobiles et de construction, en particulier dans le segment du service après-vente. Les organisations de services financiers qui essaient d'optimiser les investissements et les mouvements de marché profitent aussi largement de l'intelligence artificielle et de l'apprentissage automatique pour comprendre et gérer de multiples scénarios. Par ailleurs, les industries qui misent sur l'IoT (acteurs des secteurs de fabrication, détail, service et de l'énergie) modernisent et transforment leurs charges de travail pour optimiser les coûts et ressources et pour appliquer les analyses aux données de capteurs afin d'exécuter des modèles d'apprentissage automatique.

IDC voit aussi le RGPD comme un élément déclencheur qui a poussé les organisations à évaluer le paysage de leurs données et à augmenter leur investissement en gestion des données. Avec cette norme, la confidentialité et la gestion des données sont devenues primordiales et représentent un investissement clé des entreprises.

IDC croit que les organisations qui utilisent le RGPD comme une opportunité de sécuriser leur investissement pour développer une hygiène des données et bâtir leur stratégie d'analyse, comme la banque Barclays, sont les plus concurrentielles. Une plate-forme de données moderne basée sur la sécurité, hébergée dans des environnements régionaux hyperéchelle offre aux entreprises la capacité de se conformer aux normes tout en profitant des fonctionnalités du cloud public. Le sondage d'IDC sur le RGPD a révélé que l'appétit des entreprises pour le cloud ne diminue pas mais devient stratégique et bien calculé pour surmonter les risques de conformité au règlement. Nous pensons que les plates-formes d'analyse des données du multicloud peuvent servir de fondement aux initiatives modernes d'IoT, d'apprentissage automatique et d'intelligence artificielle, et aux nouvelles charges de travail.

La fonctionnalité SDX de Cloudera aide les entreprises à unifier, sécuriser et gouverner les données à travers de multiples silos, simplifiant ainsi la gestion de la conformité et de la confidentialité.

La gestion du chiffrement et des métadonnées est très importante dans la plate-forme de Cloudera, accompagnée des fonctionnalités de sécurité d'Azure et de stratégies de données, le tout sans compromettre la confidentialité, ce qui aide à respecter des clauses spécifiques du RGPD et offre aux entreprises européennes la confiance nécessaire pour exécuter dans leur cloud.

IDC observe comment beaucoup d'acteurs du secteur automobile en Europe voient les utilisations d'échange de données comme de la maintenance prédictive pour optimiser leurs cas commerciaux. Les fabricants de voitures utilisent de plus en plus

l'analyse des données et passent de l'échelle téraoctet à pétaoctet. D'autres utilisations comme l'optimisation des biens, la répartition des investissements et l'optimisation du réseau émergent rapidement et en force avec les analyses basées sur l'intelligence artificielle. La combinaison de l'optimisation de réseau interne et externe rencontre aussi du succès, comme vu avec Deutsche Telekom. IDC observe aussi le besoin croissant d'analyses combinées d'ensembles de données internes et externes (en traitant les données à travers tout l'écosystème), dans les secteurs de détail et de la fabrication, pour améliorer la gestion logistique, prédire l'heure prévue d'arrivée (HPA) et la rendre plus utile à la logistique.

IDC croit que les plates-formes unifiées d'analyses de données comme Cloudera sur Azure, simplifient les analyses et élargissent les utilisations. Beaucoup d'organisations rassemblent des dispositifs d'entrepôts de données pour supporter leurs exigences complexes d'analyses comme la surveillance du marché et les analyses de conformité, mais les résultats sont disparates à cause de la complexité de l'intégration. L'évolutivité changeante des solutions d'entrepôts a aussi obligé les entreprises à stocker les données par marché.

Pourtant, la consolidation de tous les ensembles de données dans une seule plate-forme d'analyses, l'évolutivité et les fonctionnalités sur demande du cloud, les fonctionnalités de gestion de la sécurité et des données et une intelligence artificielle moderne peuvent apporter beaucoup de valeur aux stratégies analytiques.

Éliminer les silos et exécuter des requêtes approfondies à échelle sont des considérations importantes à prendre en compte pour les entreprises européennes quand elles investissent dans des stratégies de transformation des données.

Perspectivas futuras

Selon IDC, les deux vendeurs travaillent de concert pour cette alliance, ce qui rend leur offre très attractive. Par exemple, sur la feuille de route, les deux prévoient d'explorer les conteneurs, l'informatique en périphérie et l'intelligence artificielle dans le cloud, et les services de gestion d'analyse des données, pour un nouveau groupe de clients comme les PME.

Bien que les vendeurs offrent une assistance et des conseils techniques et des preuves de concept communes pour aider les clients à adopter et utiliser efficacement ces technologies, Microsoft et Cloudera doivent réfléchir sur la manière de surmonter le manque de compétences, à travers des formations partenaires, l'élaboration de meilleures pratiques et l'échange de preuves de valeur pour illustrer ce que la technologie apporte au client en terme de revenus commerciaux. Les vendeurs doivent aussi être conscients de la concurrence croissante de la part d'autres acteurs de l'hyperéchelle et doivent innover plus rapidement puisqu'on se trouve maintenant à un tournant critique : les clients passent des prototypes aux environnements de production.

Miser sur les données et surpasser la concurrence à l'ère numérique

Nous entrons dans une ère de mesures de capteurs, de calculs et de prises de décisions, pour transformer les données en valeur ajoutée. On peut affirmer que les entreprises se convertissent en entreprises de l'information. Les données et informations représentent une opportunité unique de créer de la valeur jusque-là inimaginable.

Les entreprises qui souhaitent se différencier dans l'économie numérique doivent construire leur noyau intelligent au cœur d'une plate-forme numérique. IDC prévoit que d'ici 2020, 80% des entreprises auront des fonctionnalités de gestion et de monétisation des données, augmentant ainsi les fonctions de l'entreprise, renforçant la concurrence et créant de nouvelles sources de revenus. De plus, d'ici 2023, 95% des entités auront incorporé de nouveaux ensembles numériques de KPI et seront tournées vers les taux d'innovation de produit/service, la capitalisation des données et l'expérience du personnel.

Les utilisateurs finaux doivent être conscients que la technologie ne saura le faire seule. Ils doivent identifier les problèmes clés à résoudre, développer une culture basée sur les données et bien connaître les caractéristiques des charges de travail afin d'associer la bonne charge de travail à la bonne plate-forme analytique et d'utiliser les fonctionnalités adéquates.

Avec l'énorme innovation du secteur de l'entreposage des données, l'hyperéchelle et la gestion des données dans le cloud, IDC pense qu'il est temps aujourd'hui de mettre en place les fondements techniques pour pouvoir miser sur les données à l'ère numérique.

Alors que les entreprises le font, elles devraient évaluer les options technologiques comme la plate-forme de données de Cloudera sur Microsoft Azure.

IDC UK

5th Floor, Ealing Cross,
85 Uxbridge Road
Londres
W5 5TH, Royaume Uni
44.208.987.7100
Twitter: @IDC
idc-community.com
www.idc.com

Copyright y restricciones:

Toute information ou référence concernant IDC à des fins publicitaires, journalistiques ou promotionnelles nécessitent un accord préalable écrit de la part d'IDC. Pour les demandes d'autorisation, veuillez contacter le service d'information client au +1-508-988-7610 ou à l'adresse : permissions@idc.com.

La traduction et/ou localisation de ce document nécessite une licence complémentaire de la part d'IDC.

Pour plus d'informations sur IDC, aller sur www.idc.com.

Pour plus d'information sur les solutions client d'IDC, aller sur http://www.idc.com/prodserv/custom_solutions/index.jsp.

Siège social : 5 Speen Street
Framingham, MA 01701 États-Unis.

Tél. +508.872.8200
Fax. +508.935.4015
www.idc.com.

Copyright 2019 IDC.

Reproduction interdite sauf autorisation préalable

Tous droits réservés.

À propos d'IDC

International Data Corporation (IDC) est le premier groupe mondial de conseil et d'études sur les marchés des technologies de l'information, des télécommunications et des technologies de consommation. IDC aide les professionnels de l'informatique, les directeurs d'entreprises et les investisseurs à prendre des décisions éclairées concernant leurs achats technologiques et leur stratégie commerciale.

Plus de 1100 analystes, répartis dans plus de 110 pays à travers le monde, offrent une expertise locale, régionale et mondiale sur les opportunités et tendances technologiques et industrielles. Depuis 50 ans, IDC offre des informations stratégiques pour aider ses clients à atteindre leurs principaux objectifs commerciaux. IDC est une filiale d'IDG, plus grande entreprise mondiale de communication, recherche et événements liés à la technologie.