Ovum
TMT intelligence | informa

Maîtriser le cloud pour exploiter au mieux le Big Data

Un livre blanc Ovum pour Cloudera et Microsoft

Date de publication : Le 15 juin 2018

Auteur : Tony Baer

Résumé

Impulsion

Aujourd'hui, les organisations sont confrontées à une diversité de défis, de l'engagement de la clientèle à la gestion du risque, en passant par la cybersécurité, la détection des fraudes et l'excellence opérationnelle. Il est donc urgent pour elles de devenir orientées données. De plus en plus, le cloud joue un rôle clé pour aider les entreprises à mettre les données à profit, car il permet d'éviter une bonne part des coûts et des complications organisationnelles associés à la mise en œuvre de nouvelles capacités dans des centres de données physiques. Les résultats sont particulièrement probants dans le domaine du Big Data. Une grande part des charges de travail associées au Big Data sont en cours de migration vers le cloud. Selon Ovum, 27,5 % des charges de travail liées au Big Data sont exploitées dans le cloud et, à l'heure actuelle, ce taux augmente de plus de 20 % par an. Ovum prévoit que plus de la moitié des *nouvelles* charges de travail liées au Big Data seront réalisées via le cloud d'ici 2019. Pour la plupart des organisations, un déploiement hybride employant un centre de données et le cloud deviendra la norme. Si peu d'entre elles sont susceptibles de migrer 100 % de leurs données, applications et plateformes vers le cloud, celui-ci jouera un rôle absolument essentiel en matière d'analytique et d'agilité commerciale – dépassant le cadre restreint du testing/développement pour rapidement accueillir de nouvelles charges de travail. Mais attention : tous les services cloud ne sont pas les mêmes. Comment les entreprises peuvent-elles choisir le bon fournisseur d'infrastructure cloud (IaaS) *et* la meilleure plateforme Big Data à exploiter dans cet environnement cloud ? Ces décisions doivent être envisagées séparément.

La position d'Ovum

Devenir orienté données implique la capacité de desservir toutes les parties prenantes qui travaillent avec des données. Chacune d'entre elles a ses propres exigences. Les scientifiques de données requièrent un environnement permettant de facilement développer des modèles et capable de traiter l'ensemble des données d'une grappe, et non seulement un petit échantillon sur un ordinateur portable. Pour les analystes commerciaux, l'environnement doit offrir une capacité d'analyse en libre-service. Les ingénieurs de données, eux, ont besoin d'un environnement rentable et hautement performant où ils peuvent nettoyer, intégrer et transformer les données. Bâtir des systèmes dédiés pour satisfaire à ces différents besoins, cependant, est susceptible d'engendrer de nouveaux silos de données et de créer une situation non viable à long terme. Les entreprises ont donc besoin d'un environnement souple, capable de prendre en charge chaque groupe d'utilisateurs, et où non seulement les données, mais aussi les politiques de gouvernance et de sécurité, sont partagées et appliquées de manière homogène et systématique. De nombreuses plateformes cloud offrent la gamme de services requise, mais souvent avec des fonctions de sécurité périmétriques qui n'intègrent pas la gouvernance sur l'ensemble des diverses plateformes de données. Pour Ovum, un *portefeuille unifié* de services *cloud gérés*, optimisés pour une vaste gamme de charges de travail et hautement compatibles avec les centres de données hybrides sur site, constituera un outil clé pour permettre aux organisations d'accélérer le déploiement de solutions viables pour toutes les parties prenantes, à partir d'une base commune.

Messages clés

- Pour faire face aux défis familiers que représentent l'engagement de la clientèle, la réduction du risque et une meilleure excellence opérationnelle, les entreprises nécessitent des solutions à même d'être mises en œuvre rapidement et rentablement.
- Le cloud permet de significativement réduire les délais pour tirer un bénéfice commercial concret des données.
- Les services cloud gérés qui automatisent l'approvisionnement, l'exploitation et les processus correctifs seront essentiels pour les organisations souhaitant maximiser le potentiel du cloud, notamment en termes de délais de rendement et d'agilité.
- Le partenariat Cloudera/Microsoft démontre les avantages d'une PaaS (plateforme en tant que service) de Big Data gérée quand celle-ci est optimisée pour la plateforme cloud sous-jacente. En effet, cette plateforme conjugue les avantages du cloud en termes de coûts et d'agilité avec un stockage cloud évolutif et optimisé pour les problèmes de calcul complexes associés à l'analyse du Big Data, à l'ingénierie de données et à l'apprentissage machine.

Les services cloud gérés exploitent le vrai potentiel du cloud

Les défis commerciaux et organisationnels

Les défis auxquels sont confrontées les entreprises sont bien connus. Parmi les principaux :

- comment capter l'empreinte numérique des clients pour en tirer des connaissances exploitables commercialement ?
- comment connecter les produits et services en tirant parti des données et en utilisant l'internet des objets (IoT) à la périphérie ?
- comment se protéger contre les cybermenaces et la fraude ?

Le *grand défi*, c'est que la solution à tous ces problèmes doit satisfaire aux besoins de plusieurs catégories de parties prenantes/rôles au sein de l'organisation. Elle doit convenir aux scientifiques de données, à qui il faut un environnement productif pour expérimenter avec les données et les modèles. Cet environnement doit offrir une sélection exhaustive de langages, d'outils et de cadres pour la création et l'essai de modèles ; en outre, il doit permettre aux scientifiques de données de passer sans accroc d'un petit modèle sur un ordinateur portable à l'exécution sur la grappe entière. Ensuite, il y a la nécessité de mobiliser les données. Ce sont les ingénieurs de données qui s'en chargent, et eux ont besoin d'un environnement rentable et évolutif pour transformer les données. Enfin, n'oublions pas les parties prenantes du côté affaires – après tout, tous ces problèmes relèvent en fin de compte de leur responsabilité, et ce sont elles qui détiennent le pouvoir d'approbation final sur les solutions. Les analystes commerciaux exigent un environnement productif conçu pour une utilisation en libre-service, où ils peuvent exploiter les outils de veille commerciale qui leur sont déjà familiers.

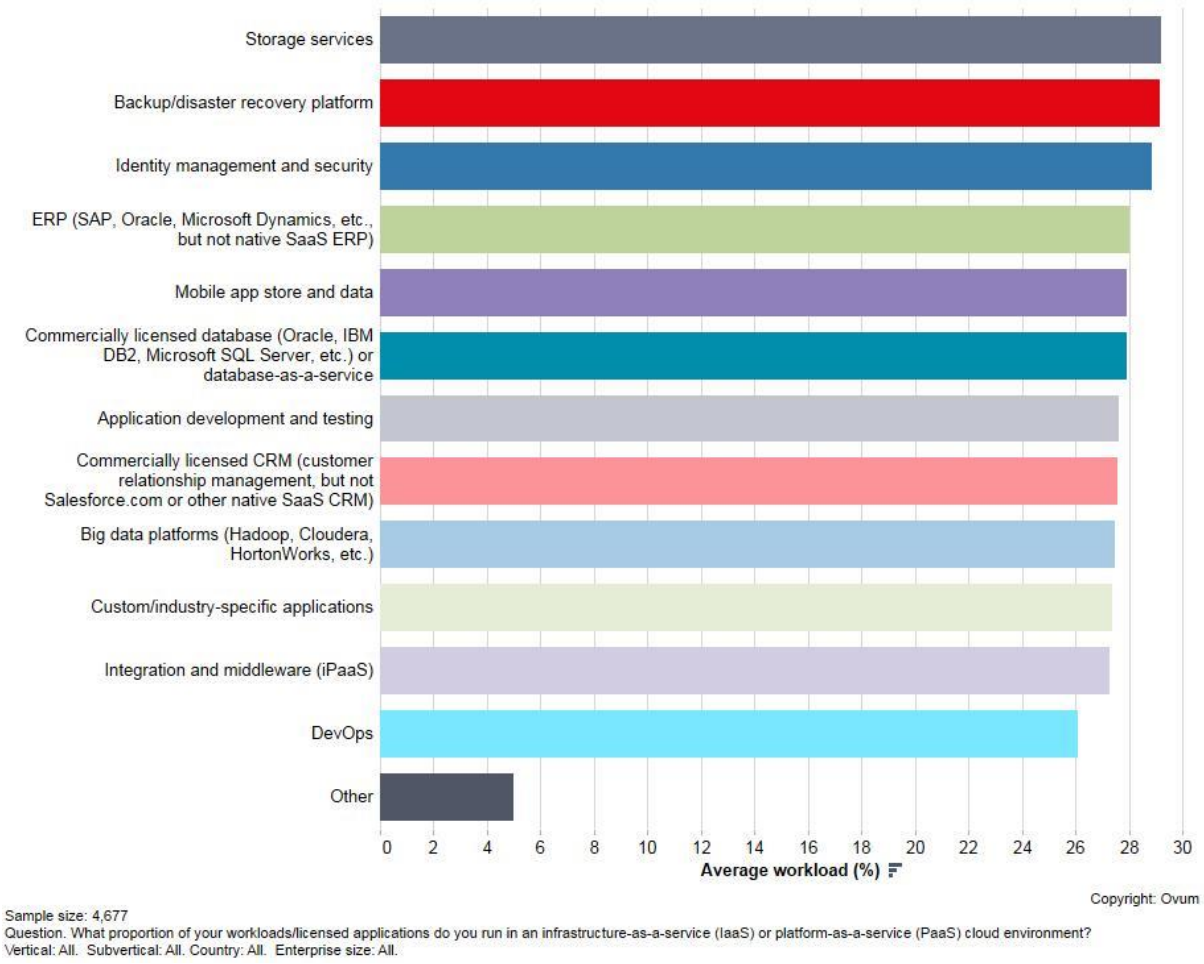
L'*opportunité* réside dans une solution parfaitement adaptée à tous ces besoins. Cependant, ce n'est pas une mince affaire, car les charges de travail que demandent les scientifiques de données, les

ingénieurs de données et les analystes commerciaux sont susceptibles d'être bien différentes. Les grappes, en effet, doivent être configurées différemment. Par exemple :

- Il est essentiel pour les scientifiques de données que leur expérience soit homogène, et qu'ils puissent d'abord concevoir des expériences sur un ordinateur portable, avant de les déployer à plus grande échelle sur la grappe entière. Les grappes qu'ils utilisent peuvent nécessiter une capacité de calcul puissante ou dense et, dans certains cas, un matériel spécialisé, comme des GPU pouvant traiter des modèles d'intelligence artificielle à l'aide de l'apprentissage machine ou de l'apprentissage profond.
- Les ingénieurs de données cherchent à moderniser la transformation des données, en tirant parti de l'infrastructure en tant que commodité, de logiciels open source et de moteurs de calcul distribués optimisés, tels que Spark. En termes de calcul, leurs besoins sont très axés IOPS.
- Les analystes commerciaux exigent un environnement efficace pour les requêtes interactives afin de permettre l'utilisation en libre-service du Big Data. Pour les organisations dans lesquelles la population d'utilisateurs est importante du côté affaires, des grappes optimisées peuvent être nécessaires pour une plus grande congruence.

Bien que les divers groupes de parties prenantes aient des besoins divergents, il y a aussi des points communs importants. La fonction informatique et la fonction affaires exigent toutes deux un environnement unifié, sécurisé et à gouvernance systématique, qui protège contre les cybermenaces, permet d'assurer la confidentialité des données et répond aux mandats réglementaires tels que le RGPD (Règlement général sur la protection des données), qui vient d'entrer en vigueur pour les organisations actives dans l'UE. Chaque fonction ne peut pas passer son temps à gérer et gouverner de multiples silos de données, chacun avec ses propres capacités (et limites).

Figure 1 : Part des charges de travail exploitées dans le cloud



Source : Ovum ICT Enterprise Insights

Pourquoi utiliser le cloud ?

Aujourd'hui, de plus en plus d'entreprises adoptent le cloud. Selon les études de veille économique d'Ovum, environ 25 à 30 % des charges de travail des entreprises sont actuellement exploitées dans le cloud ; le Big Data, figure fermement dans cette plage, à 27,5 % (cf. Figure 1). En ce qui concerne les *nouvelles* charges de travail Big Data, Ovum s'attend à ce que le cloud, qui en exécutera plus de la moitié, fasse figure de consensus d'ici l'année prochaine. Le cloud aide les entreprises à surmonter une bonne part des obstacles associés au déploiement du Big Data, à commencer par l'investissement et le temps requis pour que les équipes informatiques puissent évaluer, procurer, installer et intégrer de nouvelles capacités de calcul. En permettant aux organisations d'éviter les freins liés à un approvisionnement et un déploiement sur site, le cloud offre une plus grande agilité. Pour les clients, l'ajout de nouvelles capacités – telles que l'utilisation de l'apprentissage machine, la modernisation de l'ETC, et le lancement du libre-service – est simplement une question de réserver une certaine capacité cloud.

En outre, les déploiements basés sur une architecture native au cloud, où le stockage est séparé du calcul, fournissent des avantages complémentaires. Les clients dont la charge de travail varie beaucoup et/ou susceptibles d'expérimenter des pics soudains profitent de *l'élasticité* du cloud pour

accroître et réduire à la demande la puissance de calcul. Cela crée une souplesse considérable : au lieu d'acheter une capacité basée sur les charges de travail anticipées, les clients ne paient que la capacité de calcul dont ils ont réellement besoin.

Le monde sera hybride

Dans la plupart des organisations, les déploiements sur site ne disparaîtront pas. Par exemple, les entrepôts de données opérationnels de longue date/persistants, qui livrent des rapports à intervalles programmés, continuent souvent d'être exploités sur site, tandis que le cloud s'adapte plus facilement aux nouvelles applications ou aux charges de travail transitoires.

En matière de répartition des charges de travail entre les serveurs physiques et le cloud, il ne faut pas s'attendre à une solution miracle ou à une recette unique. Chaque organisation aura des critères propres, en fonction notamment des politiques internes ou des mandats externes sur l'emplacement physique des données stockées, de la disponibilité d'une application ou d'un système spécifique dans le cloud, de la capacité disponible pour la prise en charge d'une charge de travail spécifique, des réglementations gouvernementales ou sectorielles et des meilleures pratiques, ainsi que d'autres critères. Il est préférable qu'une entreprise maintienne une liberté de choix quant à l'exploitation de ses charges de travail, et que son fournisseur cloud soit capable de prendre en charge une telle configuration hybride, de manière transparente et uniforme à travers tous les environnements.

Types de services cloud

Il n'y pas de service cloud générique – le cloud offre une vaste étendue d'options.

Infrastructure en tant que service (IaaS) et Logiciel en tant que service (SaaS)

L'IaaS représente le service le plus basique. Avec l'IaaS, l'organisation prend un abonnement à une infrastructure de calcul brut et de stockage, soit via un contrat à long terme, soit par répartition (« pay-as-you-go »). La fonction informatique a les mêmes responsabilités, en termes de gestion des opérations, qu'elle aurait avec ses propres centres de données : sélection, approvisionnement et gestion de l'infrastructure, ainsi qu'installation, mise à jour et correction des logiciels.

En revanche, si l'organisation adopte le SaaS, les applications sont hébergées et gérées par le prestataire dans le cloud. Le client prend un abonnement pour l'application et ne gère ni l'infrastructure ni le déploiement, lesquels sont la responsabilité du fournisseur d'application.

Plateforme gérée en tant que service (PaaS)

Le PaaS représente une autre catégorie importante de services cloud. Ovum classifie les services PaaS comme services cloud « gérés », car c'est le fournisseur de *plateforme*, dans ce cas de figure, qui gère activement le service cloud, avec une approche prescriptive. Cette approche prescriptive implique la livraison d'un ensemble préconfiguré de services cloud, conçu pour réduire ou éliminer le besoin pour le client de spécifier et de gérer l'infrastructure et les logiciels. Les services cloud gérés qu'offrent les fournisseurs de plateformes peuvent comprendre des bases de données, le développement d'environnements d'applications, des services d'intégration, et d'autres outils.

Les services cloud gérés offrent la voie de déploiement la plus simple et la plus rapide

Pour Ovum, les services cloud PaaS deviendront la solution de référence pour la prochaine vague d'organisations adoptant le Big Data. Notre conviction se fonde sur le fait que ces services permettent au client de se concentrer strictement sur les données et l'analyse, ce qui accélère le délai de rentabilisation. Cela sera particulièrement vrai pour les nouveaux clients qui ont peu (ou n'ont pas) d'expérience en matière d'analytique Big Data.

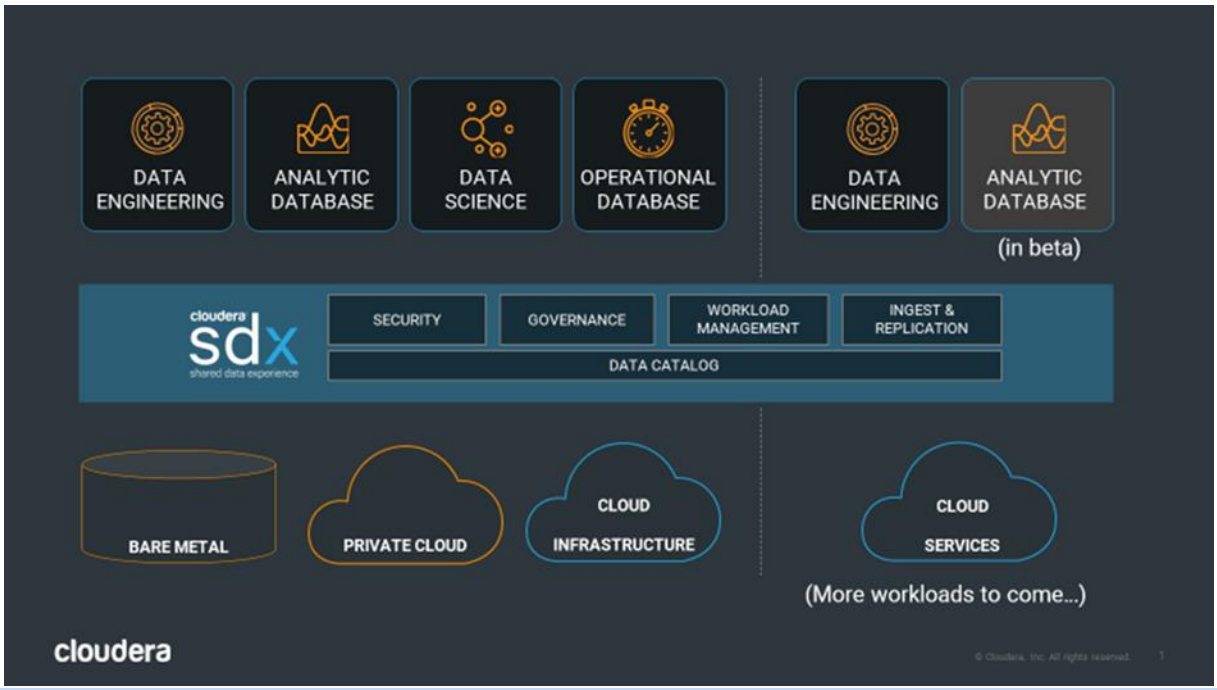
De plus en plus, les services PaaS sont conçus et commercialisés autour de charges de travail spécifiques, offrant ainsi des services optimisés en fonction du type d'utilisateur/rôle et de charge de travail. Le besoin de configurer les modalités du cloud pour tenir compte des variations dans les exigences de traitement et de stockage associées à différents types d'utilisation, comme l'analyse en libre-service, la modélisation IA, la science de données et la transformation des données, est de ce fait éliminé.

Comment le partenariat Cloudera/Microsoft répond aux besoins cloud des clients

La plateforme Cloudera est axée sur le choix

Conçue comme plateforme d'apprentissage machine et d'analytique, Cloudera Enterprise offre un choix complet aux clients en matière de fonctionnalité et de déploiement. Elle peut être déployée sur site en mode « bare metal » (métal nu) à locataire unique, dans un cloud privé, ou dans le cloud public de votre choix. Grâce à une gamme complète d'éditions, les clients de Cloudera peuvent obtenir la configuration qui convient le mieux à leurs exigences commerciales/d'utilisation. L'édition d'entreprise est conçue pour prendre en charge plusieurs types de charges de travail, et il y a des éditions spécialement configurées pour l'ingénierie de données, la science de données, et les bases de données opérationnelles. Toutes les éditions de Cloudera Enterprise sont gérées et gouvernées via une SDX (Shared Data Experience) commune, garantissant l'uniformité aux niveaux de la sécurité, de la gouvernance, de la gestion des charges de travail et de l'ingestion/réplication. La totalité des fonctions et services gouvernés par la SDX sont dirigés par un catalogue de données partagé.

Figure 2 : Options de déploiement Cloudera



Source : Cloudera

Services PaaS sur Azure, personnalisés en fonction de vos charges de travail

Cloudera et Microsoft ont formé un partenariat pour optimiser le déploiement de toutes les éditions Cloudera à partir du cloud Azure, via une architecture élastique qui sépare le calcul du stockage. Cela permet aux clients de Cloudera de tirer pleinement parti des avantages économiques d'une capacité de calcul basée dans le cloud.

Cloudera Enterprise sur Azure : une plateforme généralisée pour l'exploitation de charges de travail multiples

Cloudera Enterprise est disponible sur le cloud Azure. Elle offre plus de 100 services, une vaste sélection d'outils de bout en bout, et des accords de niveau de service (SLA) appuyés financièrement, avec des garanties de temps de disponibilité supérieur à 99 %. Disponible sur Azure Marketplace, Cloudera Enterprise sur Azure peut être déployée en un seul clic. Cela élimine les délais associés à l'achat et l'installation de nouveaux nœuds dans un centre de données physique.

Cloudera Enterprise sur Azure prend en charge une vaste plage de charges de travail, dont Impala, HBase, Spark, et Solr. Elle s'intègre aux plateformes Microsoft de données et d'analytique, réquisitionnant des données structurées et non structurées auprès de SQL Server via PolyBase et exploitant Impala pour ouvrir l'accès aux analystes commerciaux via Power BI.

Cloudera Altus sur Azure : une solution conçue conjointement pour des charges de travail spécifiques

En incorporant Cloudera Altus, la famille de services PaaS gérés de Cloudera, le partenariat Cloudera/Microsoft a franchi une nouvelle étape. Le lancement d'Altus, qui est optimisé pour Azure,

est l'œuvre conjointe de Cloudera et Microsoft. Altus fournit un service géré natif au cloud qui maintient l'élasticité de la capacité de calcul, maximisant la valeur pour les clients dont les charges de travail sont transitoires ou hautement variables.

Sous le capot, Cloudera Altus sur Azure exploite ADLS (Azure Data Lake Store), un référentiel hors catégorie pour les charges de travail d'analyse Big Data à l'échelle de l'entreprise. Le lac de données Azure permet de recueillir des données de toute taille, de tout type et de toute vitesse d'ingestion en un seul endroit à des fins d'analyses opérationnelles et exploratoires. Il fournit une durabilité extrêmement robuste, un facteur particulièrement critique pour les charges de calcul complexes et itératives – comme l'ingénierie des données – qui utilisent Spark.

Cloudera Altus for Data Engineering est le tout premier service Altus lancé sur le cloud Microsoft Azure. Cloudera Data Engineering est une charge de travail fondamentale pour le développement et l'exploitation de pipelines de données à des fins de transformation de données et de formation des modèles d'apprentissage machine. En tant que telle, l'ingénierie des données peut être considérée comme la rampe de lancement de toute initiative d'apprentissage machine et d'analyse Big Data. Exploitée en mode locataire unique, Altus met à profit la gouvernance SDX, y compris le catalogue de données partagé qui est au cœur de la plateforme Cloudera. Ovum s'attend à ce que Cloudera élargisse le portefeuille Altus dans le cloud Azure et que la société mobilise la gouvernance SDX commune dans le cadre de ce processus d'élargissement.

Points clés

Aujourd'hui, de plus en plus d'entreprises adoptent le cloud – notamment pour les charges de travail associées à l'analytique Big Data, à la science des données et à l'apprentissage machine. Le cloud permet de significativement réduire les délais pour tirer un bénéfice commercial concret des données. Les entreprises ont besoin de modalités de déploiement rapides et rentables pour faire face à leurs défis commerciaux. Les services cloud gérés seront essentiels pour les organisations souhaitant maximiser le potentiel du cloud, particulièrement en termes de délais de rendement et d'agilité. Le partenariat Cloudera/Microsoft livre les services Altus sur Azure, optimisés pour l'infrastructure Azure et intégrant les plateformes de données et les outils d'analyse de veille économique de Microsoft ; en termes de délais de rentabilité pour les clients, on peut dire que 1 + 1 font 3.

Annexe

Auteur

Tony Baer, analyste principal, gestion des données

tony.baer@ovum.com

Ovum Consulting

Nous espérons que cette analyse vous aidera à prendre des décisions d'affaires informées et imaginatives. S'il vous faut des informations complémentaires, l'équipe de conseil spécialisé d'Ovum peut sûrement vous aider. Pour en savoir plus sur les capacités de conseil spécialisé d'Ovum, veuillez nous contacter directement à l'adresse consulting@ovum.com.

Avis de droit d’auteur et de non-responsabilité

NOUS CONTACTER

www.ovum.com
analystsupport@ovum.com

BUREAUX INTERNATIONAUX

Pékin
Dubaï
Hong Kong
Hyderabad
Johannesburg
Londres
Melbourne
New York
San Francisco
São Paulo
Tokyo