

InDetail Article de Bloor
Auteur **Philip Howard**
Date de publication **Mai 2018**

IBM Cloud Private for Data



“

Nous voudrions particulièrement souligner que l'une des difficultés rencontrées par les entreprises qui tentent de mettre en œuvre l'IA est un décalage entre les spécialistes des données qui développent les modèles pertinents et ceux chargés de les déployer en production.

”

Synthèse

Intelligence artificielle (IA) et apprentissage automatique (un sous-ensemble de l'IA) sont les sujets d'actualité au sein des communautés spécialisées dans l'informatique et l'analyse, mais également dans les entreprises de façon générale. Il y est constamment fait référence dans les médias et dans les conseils d'administration. Il est généralement admis que la mise en œuvre des techniques et technologies associées à l'IA et à l'apprentissage automatique présente un potentiel énorme, à la fois directement, en termes d'efficacité accrue et indirectement, en tant que facteur de différenciation face à la concurrence. Sans parler de la possibilité d'introduire de nouveaux modèles métier et de nouveaux services. Néanmoins, s'il est facile de voir les avantages potentiels, leur mise en œuvre n'est pas aussi simple. Les prévisions suggèrent – voir **Figure 1** – qu'un très grand nombre d'entreprises vont investir dans l'IA dans les années à venir, mais aujourd'hui, seule une poignée d'entre elles ont pu le faire. Plusieurs raisons à cela : problèmes de changement culturel, problèmes de sécurité, acquisition de talents, etc., raisons que nous ne sommes pas en mesure de discuter. Cependant, il existe un problème spécifique – une capacité technologique limitée ou inexistante en ce qui concerne les données et l'analyse – que nous allons examiner dans le présent document avec IBM Cloud Private for Data.

IBM Cloud Private for Data est une plateforme intégrée de Data Science, d'ingénierie des données et de construction d'applications analytiques, bâtie au-dessus d'IBM Cloud Private (ICP). ICP est destiné à : a) fournir tous les avantages du cloud computing, mais à l'intérieur de votre pare-feu, et b) à servir de tremplin, si nécessaire, pour de plus vastes déploiements de cloud (public). ICP possède une architecture de type microservices qui offre des avantages supplémentaires, dont nous allons discuter plus en détail. ICP for Data a lui-même pour objectif de créer un environnement facilitant la mise en œuvre de processus et d'opérations axés sur les données et, plus particulièrement, de supporter le développement de capacités d'intelligence artificielle et d'apprentissage automatique (machine learning) et leur déploiement. Ce dernier point est important car il peut facilement exister une déconnexion entre les spécialistes des données (qui travaillent souvent pour les départements métier) et les personnes (généralement le personnel informatique) qui doivent mettre en application le travail de ces analystes.

Dans cet article, nous aborderons d'abord ICP puis ICP for Data. Cependant, il est utile de discuter brièvement de certaines raisons générales pour lesquelles vous pourriez vouloir déployer l'une ou l'autre de ces solutions. Dans le cas d'ICP, il existe plusieurs raisons qui vont au-delà de la flexibilité et de l'évolutivité offertes par le cloud computing. La première est la capacité à moderniser les applications d'entreprise en les restructurant à l'aide des microservices proposés par ICP. La deuxième est la capacité à créer des applications cloud natives qui peuvent inclure des services de cloud public, tout en conservant vos données en toute sécurité derrière votre pare-feu. La plupart des cas d'utilisation d'ICP for Data sont en réalité une extension de ces capacités. Elles permettent d'intégrer aux applications opérationnelles des fonctions décisionnelles ou d'apprentissage automatique : entraînées au sein d'ICP for Data mais déployées soit directement, soit via une application de cloud native, selon vos préférences. Et il existe d'autres cas d'utilisation : par exemple, vous pouvez facilement déployer un ou plusieurs lacs de données en utilisant ICP for Data avec un catalogue de données et une fédération de données (voir ci-dessous) couvrant tous ces domaines, afin d'éviter les silos.

“
La mise en œuvre des techniques et technologies associées à l'IA et à l'apprentissage automatique présente un potentiel énorme, à la fois directement, en termes d'efficacité accrue et indirectement, en tant que facteur de différenciation face à la concurrence.
”

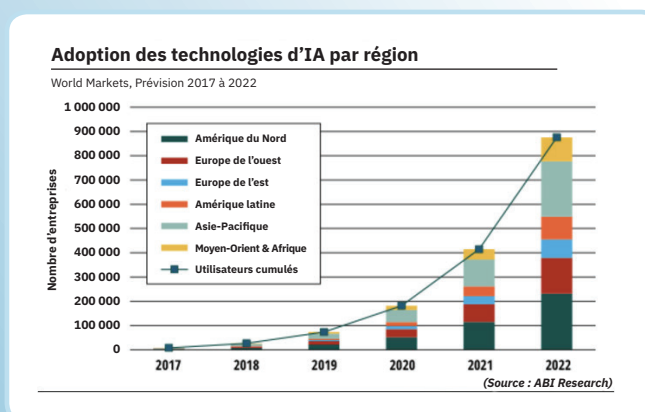


Figure 1 : Adoption prévue de l'IA

Nous n'avons cependant pas l'intention d'écrire en volume sur les caractéristiques de ces produits ni d'étudier en détails les cas d'utilisation. Notre objectif est plutôt de faire référence aux avantages métier visés par ces offres et de discuter des problèmes pouvant survenir dans cet environnement. Nous concluons par une section sur la feuille de route à court terme d'IBM pour les nouvelles versions d'ICP et d'ICP for Data.

IBM Cloud Private



Cette approche accélère considérablement les cycles de mise en production (intégration continue) et permet des améliorations incrémentielles des applications.



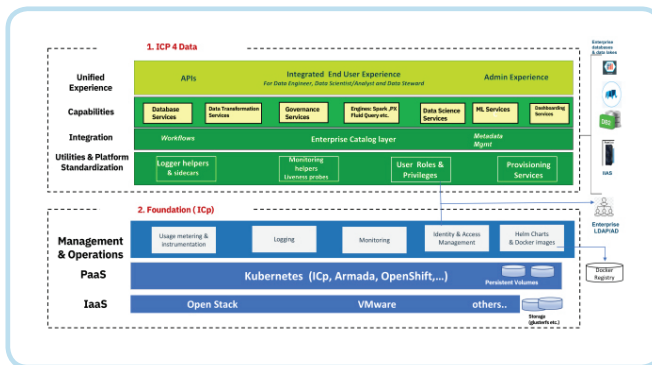
ICP est un cloud privé. Notez qu'un cloud privé est différent d'un cloud privé virtuel dans lequel un fournisseur de cloud public agit en tant que fournisseur de services. Avec ICP, votre département informatique joue le rôle de fournisseur de services ou bien vous pouvez faire appel à un tiers mais dans les deux cas, les unités métier agissent en tant que locataires. Le principe de base est que vous pouvez bénéficier des avantages d'un déploiement dans le cloud, tout en protégeant vos données derrière votre pare-feu, bien qu'elles puissent être hébergées à l'extérieur si vous le souhaitez. ICP interagit avec les déploiements traditionnels sur site, les clouds publics et les clouds publics virtuels. Il prend donc en charge les environnements hybrides et est considéré par IBM comme un tremplin potentiel vers les clouds publics du futur.

Les avantages du cloud computing, en particulier les avantages découlant de l'évolution élastique, du déploiement rapide, de l'évolutivité, etc., sont bien connus et nous n'avons pas l'intention de les répéter ici. Cependant, ICP et ICP for Data possèdent une architecture microservices, comme illustré sur la **Figure 2**, et il convient de discuter des avantages de cette architecture, car l'utilisation de microservices, basés sur Docker et Kubernetes, est un concept relativement nouveau.

Selon www.microservices.io, «les microservices, également appelés l'architecture microservice, sont un style architectural qui structure une application sous la forme d'une collection de services à couplage lâche implémentant des capacités métier. L'architecture microservices permet la distribution/le déploiement en continu d'applications volumineuses et complexes. » Microsoft va plus loin en déclarant (entre autres) « À certains égards, les microservices constituent l'évolution naturelle des architectures orientées services (SOA), mais il existe des différences. Parmi les caractéristiques qui définissent un microservice, citons le fait que les services sont petits, indépendants et faiblement couplés ; que chaque service est une base de code distincte, qui peut être gérée par une petite équipe de développement ; que les services peuvent être déployés indépendamment et qu'un service existant peut être mis à jour sans reconstruire ni redéploier l'ensemble de l'application ; et que les services communiquent entre eux à l'aide d'API bien définies dont les détails d'implémentation sont masqués pour les autres services. »

Avec ICP et ICP for Data, IBM a repris les fonctionnalités existantes et les a recréées – là où cette opération était logique – en tant que services, comme illustré sur la **Figure 2**. Certains des avantages de cette approche sont décrits dans les citations ci-dessus, mais il en existe d'autres. Par exemple, notons que cela encourage l'interopérabilité. Il est beaucoup plus facile de combiner un service de base de données avec un service d'apprentissage automatique, par exemple, car vous pouvez choisir les services que vous souhaitez utiliser. L'autre point à souligner est que cette approche accélère considérablement les cycles de mise en production (déploiement continu) et permet d'apporter aux applications des améliorations incrémentielles. C'est important, non seulement parce que vous obtenez les nouvelles fonctionnalités plus rapidement, mais aussi parce que les cycles de mise en production traditionnels d'un an ou de dix-huit mois sont généralement perturbants et entraînent des retards dans la mise en œuvre des nouvelles versions. De plus, sachant qu'il s'agit d'un environnement de cloud, IBM peut également introduire de façon incrémentielle de nouvelles fonctionnalités de la même manière.

Figure 2 : Architecture ICP for Data



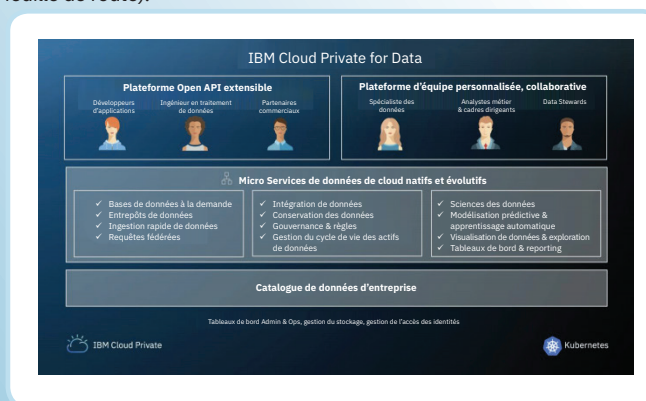
IBM Cloud Private for Data

Supporter des initiatives d'IA exige d'obtenir toutes les données dont vous avez besoin, d'en assurer la gouvernance de manière à ce qu'elles soient fiables, d'analyser et de construire l'apprentissage automatique et les autres algorithmes nécessaires au projet et, enfin, de pouvoir mettre en production les résultats de cet exercice. C'est une tâche relativement complexe. Les individus et les groupes – illustrés sur la **Figure 3**, (qui montre les composants d'ICP for Data) responsables de ces activités sont souvent disparates et déconnectés et une approche collaborative est nécessaire pour que tout fonctionne parfaitement. De plus, cette approche exige un ensemble de fonctionnalités qui dépasse les capacités de la plupart des éditeurs de logiciels. Pour permettre le type de collaboration requis, il est en outre préférable de disposer d'une interface utilisateur cohérente couvrant toutes les fonctionnalités logicielles sous-jacentes. Nous allons examiner plus en détail ICP for Data dans un instant, mais nous pouvons d'ores et déjà souligner que l'une des caractéristiques les plus remarquables de ce nouveau produit réside précisément dans le fait qu'il offre une expérience utilisateur commune pour toute sa pile logicielle. Et il est intéressant de préciser que c'est l'architecture microservices d'ICP for Data qui a permis cette approche, en séparant l'interface utilisateur des composants logiciels, puis en créant une nouvelle interface cohérente. Pour référence, la **Figure 4** fournit quelques captures d'écran illustrant cette interface.

Si nous nous reportons à nouveau à la **Figure 3** un certain nombre d'éléments de l'architecture ICP for Data méritent d'être examinés. L'élément de base est le Catalogue des données de l'entreprise. Vous pourriez le comparer à un catalogue de bibliothèque car il contient des informations sur tous les actifs de données disponibles pour l'organisation. Cependant, contrairement à une bibliothèque qui fournit simplement une référence pour chaque livre, le catalogue de données vous permet d'effectuer une recherche par catégorie et de rechercher par exemple tous les actifs liés aux ventes, aux clients ou aux produits. En d'autres termes, il vous permet de trouver des actifs présentant un intérêt pour votre rôle et dont vous ne pourriez peut-être pas avoir connaissance autrement.

Au-dessus du Catalogue des données de l'entreprise de la **Figure 3**, figurent trois cases. IBM y fait référence sous la forme « *collecter*, *organiser* et *analyser* ». S'il n'en tenait qu'à nous, nous placerions probablement l'intégration de données dans la case « *collecter* » et nous renommerions « *organiser* » en « *gouvernance* » mais ce n'est qu'une petite chicanerie sémantique.

La case « *collecter* » permet en principe de tirer parti de toutes sortes de sources de données, ce qui est évidemment nécessaire : si vous ne disposez pas des données, vous ne pouvez pas les analyser. Cependant, cet élément d'ICP for Data est relativement limité – comparé à « *organiser* » et « *analyser* » – dans cette version (il sera considérablement étendu dans la prochaine version : voir la section Feuille de route ci-dessous) et prend uniquement en charge les différents éléments de la famille Db2 (Db2, Db2 Warehouse, Db2 Event Store, etc.), ainsi qu'Hadoop (Hortonworks, avec lequel IBM a un partenariat) et Big SQL (moteur SQL on Hadoop d'IBM). Il convient de prendre en compte le fait que Db2 Event Store est supporté parce qu'il sera particulièrement important dans les environnements basés sur l'Internet des objets (IoT) dans lesquels vous devez stocker des données provenant de capteurs, d'actionneurs et d'autres dispositifs de périphérie. Également prévu dans la case « *collecter* », la fédération de données, qui autorise des requêtes distribuées sur des bases de données tierces telles qu'Oracle. Cependant, pour le moment, il s'agit d'une fédération de données entre bases de données relationnelles et non de virtualisation de données entre sources de données structurées et non structurées (voir la feuille de route).



Les capacités « *organiser* » et « *analyser* » sont plus complètes dans cette version que les services « *collecter* ». En ce qui concerne « *organiser* », ICP for Data exploite les technologies familières de la marque InfoSphere, avec support du nettoyage, du masquage, de la gouvernance des données, etc. Dans ce contexte, il convient de souligner certaines implications concernant l'IA et l'apprentissage automatique qui ne sont peut-être pas immédiatement évidentes. Considérons par exemple les prises de décision automatisées. En d'autres termes, une sorte d'algorithme informatisé prend les décisions automatiquement, dans le cadre d'un processus pertinent, en fonction

“
Certaines implications concernant l'IA et l'apprentissage automatique ne sont peut-être pas immédiatement évidentes.
”

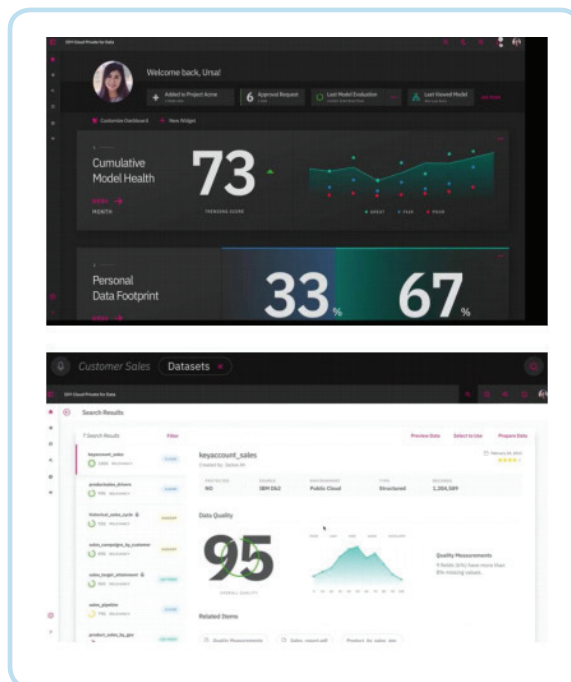
Figure 3 :
Composants ICP for Data



Ce qui est requis est « l'AnalyticOps » – comparable à DevOps – pour combler le fossé entre l'analyse et son déploiement opérationnel, ce qui est précisément ce que l'environnement de collaboration d'ICP for data est destiné à fournir.



Figure 4 :
L'expérience
utilisateur commune
d'ICP for Data



La case « analyser » de même, exploite les technologies IBM existantes, dont SPSS, Cognos et Watson Studio, qui inclut IBM Data Science Experience, produit qui offre des fonctionnalités spécifiques pour la formation, la persistance et l'évaluation des modèles d'apprentissage automatique. Cependant, il existe un fossé entre les données fiables et les modèles d'apprentissage automatique. Les données provenant de sources différentes doivent être converties en un format cohérent

et éventuellement transformées (de nombreux algorithmes fonctionnent sur des plages de valeurs comprises entre zéro et un, ce qui explique que les valeurs nulles posent problème) et doivent être reliées en un seul jeu de données avant toute analyse de ces données ou toute création d'algorithmes d'apprentissage automatique.

Nous sommes en désaccord avec IBM sur la Figure 3 en ce qui concerne la séparation entre les deux cases supérieures. Nous comprenons qu'il s'agit d'un diagramme destiné au marketing, mais il ne reflète pas exactement les intentions, ni les capacités d'IBM. Il suggère notamment que les développeurs d'applications et les ingénieurs en traitement de données ne partagent pas la « plateforme d'équipe collaborative personnalisée » – qui inclut des recommandations issues du crowd-sourcing, ainsi que des fonctionnalités plus simples telles que les workflows – disponibles pour les autres utilisateurs. Nous estimons cette présentation inexacte. Car le partage de cette plateforme devrait être général. Nous voudrions particulièrement souligner ici que l'une des difficultés rencontrées par les entreprises qui tentent de mettre en œuvre l'IA est un décalage entre les spécialistes des données qui développent les modèles pertinents et ceux chargés de les déployer en production. Et, si les ingénieurs en traitement de données sont définis comme ceux qui préparent les données pour l'analyse tandis que les spécialistes des données (data scientists) se chargent de la science, cette plateforme les concerne également. Donc, ce qui est requis est « l'AnalyticOps » – comparable à DevOps – pour combler le fossé entre l'analyse et son déploiement opérationnel, ce qui est précisément ce que l'environnement de collaboration d'ICP for data est destiné à fournir. Nous devons également ajouter que le fait de pouvoir offrir une collaboration sur tout cet éventail de rôles et de fonctionnalités n'est possible qu'en raison de l'ampleur des capacités d'IBM.

Dans ce contexte, nous devons également préciser que la gestion de modèles, bien qu'elle ne soit pas explicitement présentée dans la Figure 3, est supportée par ICP for Data. C'est un point important car le meilleur modèle d'aujourd'hui ne sera peut-être pas le meilleur modèle demain. Ceci pour deux raisons : premièrement, l'année prochaine, vous aurez davantage de données réelles avec lesquelles travailler et deuxièmement, les conditions et les tendances évoluent avec le temps. Vous devrez peut-être périodiquement remplacer un algorithme ou un modèle par un autre. Idéalement, ce changement devrait se faire sur une base de « remplacement à chaud » sans temps d'arrêt. Dans tous les cas, les performances des modèles doivent être surveillées et modifiées lorsque nécessaire, d'où la nécessité d'une gestion des modèles.

Enfin, notez qu'ICP for Data exploite toutes les capacités sous-jacentes d'ICP, en matière d'administration, de sécurité, de journalisation, de surveillance, etc.

Feuille de route

À l'heure où nous écrivons, l'édition ICP for Data Enterprise est actuellement disponible en version bêta et devrait paraître prochainement (le 29 mai 2018). Comme le montre la **Figure 5**, la Community Edition et une Cloud Native Edition, devraient sortir au cours du second semestre de 2018 (probablement plus tôt que plus tard). Cependant, d'importantes mises à niveau de l'édition Enterprise sont également prévues. Pour commencer, la première version s'exécute sous Linux sur les plates-formes x86, mais IBM a l'intention de l'étendre à d'autres plates-formes. En outre, du point de vue du cloud, ICP for Data s'exécute uniquement sur IBM SoftLayer pour le moment, mais à nouveau, IBM a l'intention de le rendre disponible sur d'autres plates-formes de cloud. Dans ce contexte, il convient de noter qu'IBM et Red Hat ont annoncé qu'ICP et ICP for Data seraient disponibles sur la plate-forme Red Hat OpenShift Container en tant que conteneurs certifiés Red Hat. IBM a également annoncé qu'ICP et ICP for Data seraient portés sur les plates-formes de cloud AWS et Azure. Enfin, IBM prévoit également de rendre ICP for Data disponible en tant qu'appliance.

En ce qui concerne les fonctionnalités individuelles, le plus notable sera l'extension des capacités du jeu de composants « collecter ». En particulier, la fédération de données sera remplacée (ou complétée) par la virtualisation des données. Et, en général, ICP for Data prendra en charge une gamme beaucoup plus large de sources de données structurées et non structurées.

Il convient de rappeler une fois de plus que l'introduction rapide des différentes fonctionnalités de la feuille de route décrites, ainsi que d'autres que nous n'avons pas mentionnées, est possible du fait la présence de l'architecture microservices. Sans elle, vous devriez attendre l'année prochaine pour obtenir ces fonctionnalités supplémentaires.

“
IBM a également annoncé qu'ICP et ICP for Data seraient portés sur les plates-formes de cloud AWS et Azure.
”

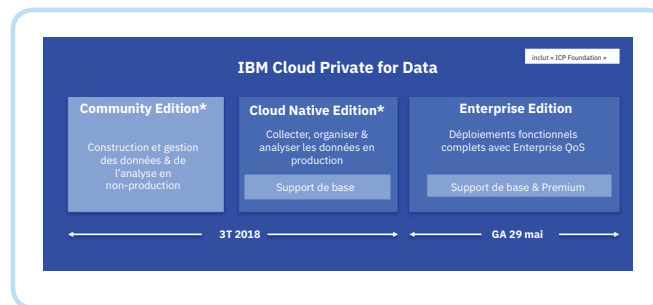


Figure 5: les éditions d'ICP for Data

Conclusion

Lors du sommet Data Works de Berlin en avril 2018 un sondage a été réalisé auprès du public, demandant combien de sociétés parmi les participants prévoient de placer des données et des analyses dans le cloud. Résultat qui peut sembler surprenant, 34 % des plus de 400 personnes ayant répondu, ont déclaré que leur entreprise n'avait pas de tels projets. La vérité est que, malgré l'ampleur du battage médiatique sur ce sujet, de nombreuses organisations hésitent encore à sauter le pas, pour diverses raisons. Cela ne signifie pas qu'elles ne reconnaissent pas les avantages du cloud computing, mais, à l'heure actuelle, il est perçu comme non encore atteignable. ICP for Data offre une option intermédiaire : les avantages du cloud computing sans le risque de déplacement des données en dehors de votre pare-feu.

Cependant, ce n'est pas tout ce que propose ICP for Data : si vous souhaitez déployer l'apprentissage automatique – et presque tout le monde le fait – vous avez besoin d'un environnement qui facilite ce déploiement. IBM y fait référence en disant que vous ne pouvez pas avoir d'intelligence artificielle sans une architecture d'information (« *L'IA exige l'AI* »). Le problème avec la construction d'une architecture d'information est qu'elle implique de nombreuses pièces mobiles, de nombreuses exigences logicielles et de nombreuses personnes. Pour que tout ceci fonctionne, les entreprises doivent adopter comme principe l'AnalyticOps. Un tel principe exige non seulement une vaste gamme de fonctionnalités de base mais également un support collaboratif pour l'ensemble des personnes concernées. Même si ICP for Data en est à sa première version, vous pouvez voir que le produit s'oriente dans cette direction. Il serait infiniment plus difficile à réaliser avec un ensemble de produits disparates provenant de plusieurs fournisseurs.



Même si ICP for Data en est à sa première version, vous pouvez voir que le produit s'oriente dans cette direction [le support d'AnalyticOps]. Il serait infiniment plus difficile à réaliser avec un ensemble de produits disparates provenant de plusieurs fournisseurs.



AUTRES INFORMATIONS

D'autres informations sur ce sujet sont disponibles sur www.BloorResearch.com/update/2397



À propos de l'auteur

PHILIP HOWARD

Directeur de recherche/Information Management

Philip a débuté dans l'industrie informatique en 1973 et a occupé divers postes en tant qu'analyste systèmes, programmeur et vendeur, ainsi que dans les domaines du marketing et de la gestion de produits pour diverses entreprises, notamment GEC Marconi, GPT, Philips Data Systems, Raytheon et NCR.

Il a ensuite créé, après un quart de siècle, sa propre entreprise en 1992 et son premier client a été Bloor Research (à l'époque ButlerBloor). Philip a travaillé pour la société en tant qu'analyste associé. Ses relations avec Bloor Research se sont poursuivies depuis et il est maintenant Directeur de recherche, spécialisé dans la gestion de l'information.

La gestion de l'information comprend tout ce qui concerne la gestion, le déplacement, la gouvernance et le stockage des données, ainsi que l'accès à ces données et leur analyse. Elle implique diverses technologies qui incluent (sans s'y limiter) les bases de données et

l'entreposage de données, l'intégration des données, la qualité des données, la gestion des données de référence, la gouvernance des données, la migration des données, la gestion des métadonnées, la préparation et l'analyse des données.

Outre les nombreux rapports rédigés par Philip pour Bloor Research, Philip contribue régulièrement à **IT-Director.com** et **IT-Analysis.com** et était auparavant éditeur de **Application Development News** et **Operating System News** pour le compte de Cambridge Market Intelligence (CMI). Il a également contribué à divers magazines et rédigé un certain nombre de rapports publiés par des sociétés telles que CMI et The Financial Times. Philip intervient régulièrement lors de conférences et autres événements en Europe et en Amérique du Nord.

En dehors du travail, ses principales activités sont les croisières sur les canaux, le ski, le bridge (niveau Life Master) et les restaurants.

Présentation de Bloor

La technologie permet l'évolution rapide des entreprises. Les opportunités sont immenses mais si vous ne vous adaptez pas, vous ne survivez pas. Donc, en cette ère de mutations, le succès repose sur l'évolution.

Nous vous montrons l'avenir et vous aidons à le concrétiser.

Bloor apporte une pensée technologique nouvelle pour vous aider à naviguer dans des situations métier complexes, transformant les défis en de réelles opportunités de croissance, de rentabilité et d'impact.

Nous fournissons des informations stratégiques exploitables par le biais de nos services de conseil innovants et indépendants dans le domaine des technologies. Nous aidons les entreprises tout au long de leur processus de transformation à rester pertinentes, en apportant une nouvelle réflexion aux situations métier complexes et en transformant les défis en de réelles opportunités de croissance et de rentabilité.

Depuis plus de 25 ans, Bloor aide les entreprises à évoluer de manière intelligente : en adoptant la technologie pour ajuster leurs stratégies et obtenir les meilleurs résultats possibles. Bloor vous aide à remettre en question les hypothèses en vue d'une amélioration et d'un succès constants.

Copyright et clause de protection

Ce document est protégé par copyright © 2018 Bloor. Aucune partie de cette publication ne peut être reproduite par quelque procédé que ce soit sans le consentement préalable de Bloor Research.

En raison de la nature de ce document, de nombreux produits matériels et logiciels ont été mentionnés nommément. Dans la majorité, voire la totalité des cas, ces noms de produits sont revendiqués en tant que marques commerciales par les sociétés qui fabriquent ces produits. Bloor Research n'a pas l'intention de revendiquer ces noms ou ces marques commerciales. De même, les logos, graphiques et captures d'écran de société ont été reproduits avec l'accord du propriétaire et sont soumis aux droits d'auteur.

Bien que la préparation de ce document ait été effectuée avec le plus grand soin pour s'assurer de l'exactitude des informations, les éditeurs ne peuvent être tenus responsables des erreurs ou omissions éventuelles.





Bloor Research International Ltd
20-22 Wenlock Road
LONDRES N1 7GU
Royaume-Uni.

Tél : **+44 (0)20 7043 9750**
Web : **www.Bloorresearch.com**
E-mail : **info@Bloor.eu**