Db2

La base de données d’IA

Sommaire

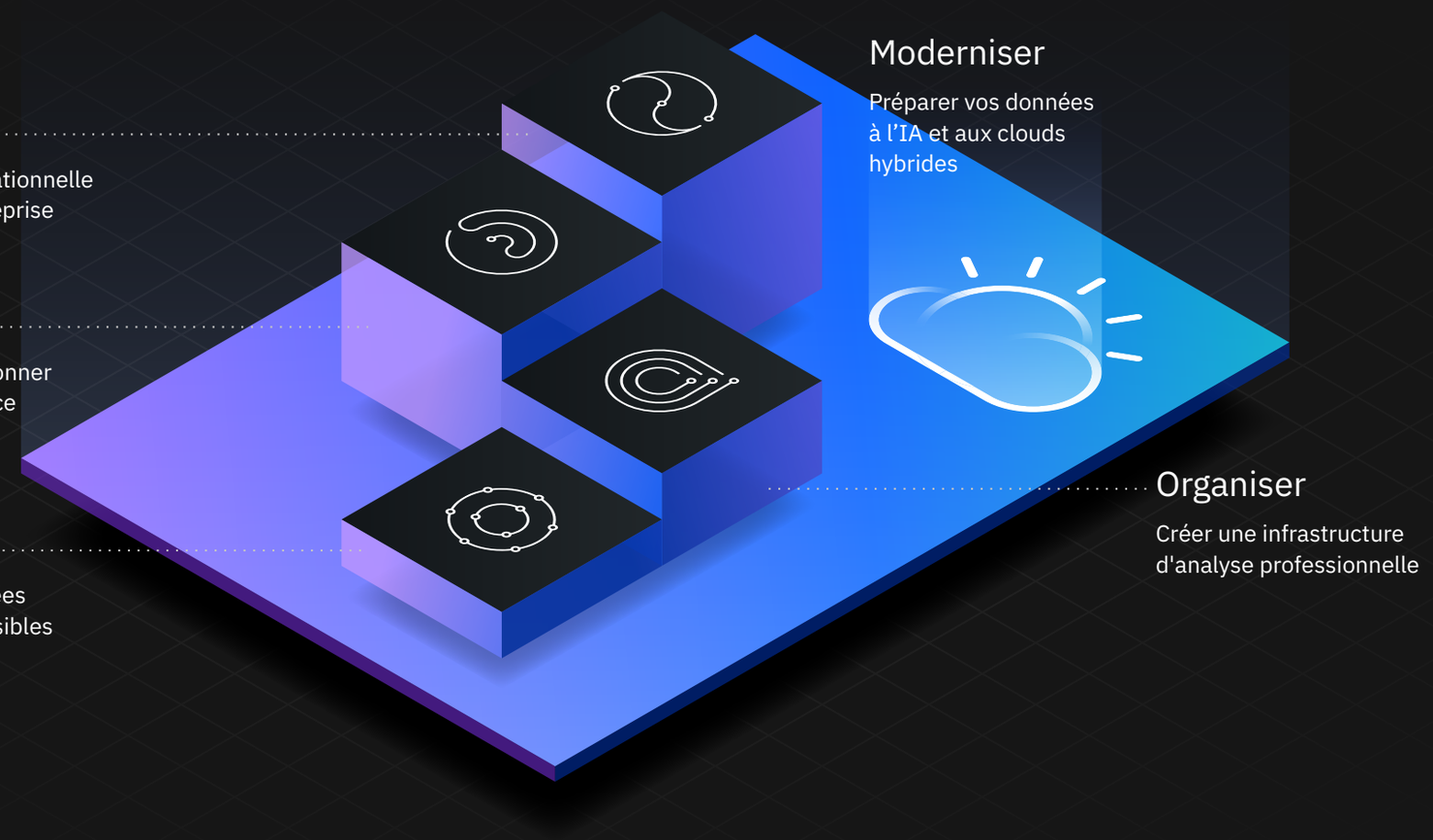
La base de données d’IA :

Prise en charge par et conçue pour l’IA

Les volumes croissants et la diversité des données exigent une nouvelle expertise, de nouveaux processus et une nouvelle infrastructure pour assurer la transformation numérique dans différents secteurs d'activité. Le Machine Learning (ML) et l’intelligence artificielle (IA) sont des composantes essentielles de ce travail, car les ordinateurs traitent les données et apprennent par expérience, sans intervention humaine continue.

L’IA et le ML facilitent les tâches informatiques traditionnelles et non traditionnelles – depuis l’analyse et la logistique jusqu’aux interactions en langage naturel, en passant par la composition de musique, la conception de produits chimiques et la détection de la fraude. Ils apportent une intelligence fondée sur les données, qui permet aux entreprises de comprendre ce qui s’est passé et de savoir ce qui pourrait advenir.

La transition vers l’IA est comme une échelle où chaque composante est un barreau. Ceux qui s’engagent dans cette transition renforcent l’échelle et la rendent encore plus utile.



La transition vers l’IA repose sur une infrastructure robuste qui permet aux organisations d’accéder aux données, de les collecter, de les organiser et de les analyser, quels que soient leur type, leur source et leur déploiement. Avec une gestion des données pilotée par le ML, l’IA est à la fois l’objectif et le moyen pour y parvenir.

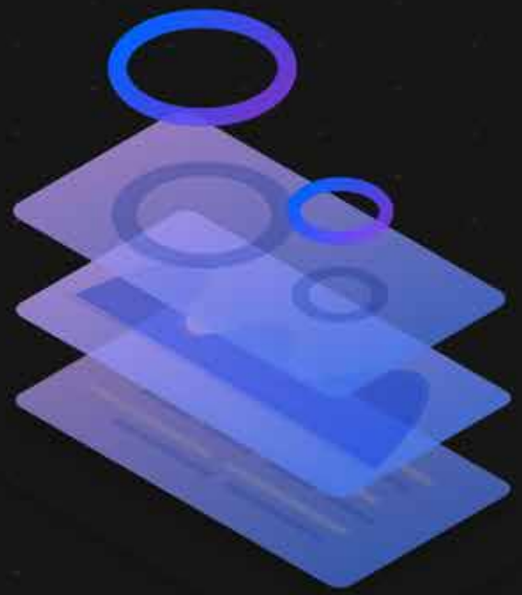
Les solutions d’IA et de ML rendent la gestion des données plus simple, plus rapide et plus intelligente, permettant aux développeurs et aux « data scientists » d’exploiter les données de nouvelles manières en intégrant les fonctionnalités d’IA à tous les niveaux de l’entreprise, depuis les opérations jusqu’à la création de modèles en passant par le développement d'applications.

Prise en charge par l’IA

Les fonctionnalités de ML et d’IA dans IBM® Db2® offrent des capacités de gestion de données inédites, qui changent radicalement la donne :

- Virtualisation complète des données
- Gestion automatique des charges de travail et optimisation des ressources
- Traitement des requêtes dix fois plus performant
- Résultats des requêtes basées sur la confiance

Virtualisation des données



Définition

C’est la combinaison de la fédération des données et d’une couche d'abstraction. Elle permet à l’ensemble des utilisateurs et des applications d’interagir avec plusieurs sources de données depuis un même point d’accès, quels que soient le type, le format, la taille et l’emplacement des données sous-jacentes. Cet accès plus large contribue à décloisonner les données et à les exploiter plus rapidement et plus efficacement pour faciliter l’adoption de l’IA.

Intérêt

Un seul point d’accès aux données offre les avantages suivants :

- **Un accès aux données simple pour vos spécialistes des données**
  Une vue unique de l’ensemble des données permet d’effectuer des recherches et d’accéder aux informations dans différentes sources de données – structurées ou non, relationnelles ou NoSQL – sans consacrer du temps et des ressources aux opérations de déplacement. Les développeurs et les ingénieurs ont accès à tous les types possibles de données – historiques et nouvelles – et peuvent extraire des connaissances de combinaisons inattendues.
- **Un seul point de contrôle pour la gouvernance et la sécurité**
  Les mesures de sécurité et de gouvernance sont bien plus simples, plus robustes et moins sujettes aux défaillances si vos administrateurs dirigent l’ensemble des utilisateurs et des applications vers un même point d'accès à tous vos référentiels de données.
- **Réduction du nombre de transferts de données**
  Les données peuvent être traitées dans les référentiels, sans avoir à les transférer d’abord dans un référentiel local. Ce qui diminue considérablement la latence et les coûts de bande passante.

Principe de fonctionnement

Les environnements d’entreprise modernes dépendent de plusieurs bases de données : données transactionnelles dans une base relationnelle, données d’opinion dans un cluster Hadoop, données d’historiques de consommation dans un entrepôt, et données sur les flux de clics du e-commerce dans un cluster rapide.

Sans la virtualisation, tous ceux qui veulent interagir avec ces données – administrateurs de bases de données (DBA), développeurs d'applications et « data scientists » – devraient mettre en place une méthode d'accès personnalisée pour chaque base de données. Au mieux, cela prendrait la forme d’un énorme casse-tête pour toutes les personnes concernées. Cela limiterait les connaissances que les utilisateurs pourraient extraire de la comparaison des différents types de données (historique client avec comportement de navigation en temps réel, par exemple). Au pire, cela pourrait endommager les données et ouvrir de nouvelles failles de sécurité.

Le fait de n'offrir qu’un seul point d’accès aux différentes données dans l’entreprise explique pourquoi la fédération des données est devenue aussi populaire. Pourtant, la virtualisation des données IBM va bien au-delà de la fédération, en fournissant une couche d’abstraction et en permettant aux utilisateurs de définir leurs propres terminologie et méthodologie concernant l’accès aux données (voir la figure 1).

En action

Virtualisation des données

La banque européenne ING travaille avec IBM à la mise en place d’un point unique d'accès aux données pour l’ensemble de ses utilisateurs à travers le monde. Elle gère ainsi une infrastructure de données globale, avec les fonctionnalités de performance, d’évolutivité, de sécurité et de gouvernance dont elle a besoin. Cela réduit également le nombre de déconnexions entre les différents sites de l’entreprise.

La solution IBM permet d'ajouter de nouvelles données à la plateforme ING depuis n’importe où dans le monde, et les utilisateurs de la banque peuvent y accéder sans qu’il soit nécessaire de modifier les droits d'accès ou les autorisations de sécurité de chaque personne.

Figure 1. Comment la virtualisation unifie l’accès à plusieurs sources de données et la gouvernance de celles-ci

Gestion adaptative des charges de travail et optimisationdesressources

Définition

Une technologie intelligente qui alloue automatiquement les ressources de données en fonction des charges de travail Cette technologie est à la fois automatisée et adaptative, car elle répond aux conditions passées et prévoit les demandes futures.

Intérêt

Avec la gestion adaptative des charges de travail, vous disposez d'un système performant, stable et prêt à l'emploi, qui ne requiert aucun ajustement ni aucune configuration nécessitant une main-d'œuvre importante. C'est un aspect essentiel pour les entreprises modernes, en raison de la complexité de leurs plateformes de données et du taux élevé de défaillance des stratégies manuelles de gestion des charges de travail.

Lorsque trop de charges de travail arrivent simultanément, une base de données doit déterminer comment gérer ce pic de demande. Ces modes manuels ou « ouverts » n'offrent aucun mécanisme de retour : ils obligent l'utilisateur à prédéfinir des limites pour le nombre ou la taille des charges de travail. Mais ces limites peuvent s'effondrer face à un jeu complexe de charges de travail, ou au mieux, nécessiter une surveillance et un ajustement constants. Ce qui entraîne une dégradation des performances, une sous-utilisation, voire une défaillance des bases de données.

En revanche, Db2 détecte et prédit les tendances d'utilisation. Il avertit les utilisateurs ou corrige automatiquement le problème avant qu'il ne prenne de l'ampleur. Cela facilite la tâche des DBA et abaisse le coût de possession de chaque base de données, en réduisant le temps de configuration et d'optimisation.
Les tests IBM ont montré que cette utilisation améliore jusqu'à 30 % les performances globales des bases de données.

Principe de fonctionnement

La gestion adaptative des charges de travail intègre le Machine Learning sous la forme d'un mécanisme de retour. Ainsi, la base de données surveille en permanence le traitement prévu et réel des différentes charges de travail, puis ajuste les ressources disponibles en conséquence.
Les bases de données Db2 sont prêtes à l'emploi, avec une configuration minimale pour la plupart des charges de travail.

Si une configuration manuelle s'impose dans les cas particulièrement complexes, cette technologie adaptative simplifie l'opération. Les DBA peuvent créer plusieurs classes de charges de travail et définir des objectifs de performance pour chacune d'elles. Ensuite, le gestionnaire adaptatif surveille les charges de travail entrantes et alloue des ressources pour atteindre ces objectifs.

En action

La gestion automatisée des charges de travail

La gestion adaptative et automatisée des charges de travail résout les principaux inconvénients liés à la gestion d'une base de données d'entreprise.

Par exemple, dans les entrepôts de données, la base de données peut recevoir des tâches très différentes : intégrations de données en flux, rapports de faible latence avec des temps de réponse inférieurs à la seconde (comme le chargement d'une page web) ou rapports groupés dont la création nécessite plusieurs minutes. L'administrateur doit s'assurer que le système ne bloque pas les tâches plus volumineuses ou les travaux urgents.

Le gestionnaire adaptatif des charges de travail simplifie tout cela, permettant aux administrateurs de créer des groupes de ressources (également appelés « classes de charges de travail ou de services ») d'allouer une part du système à chaque groupe.
Puis, la base de données alloue des ressources intelligemment pour atteindre les objectifs de performance. Le tout, sans qu'il soit nécessaire de surveiller et d'optimiser manuellement les charges de travail et les ressources individuellement.

L’IA est à la fois l’objectif et le moyen d'atteindre cetobjectif.

Optimisation des requêtes de MachineLearning

Définition

Db2 Machine Learning Optimizer offre un niveau supplémentaire d'optimisation intelligente qui utilise le Machine Learning non supervisé pour proposer des stratégies d'exécution des requêtes, plus performantes que l'optimisation traditionnelle basée sur le coût.

Intérêt

Les optimiseurs de coût de charge de travail de base peuvent proposer des stratégies d'exécution pour une requête donnée, mais comme ils ne tiennent pas compte des modifications récentes dans la base de données, ils ne peuvent pas apprendre par l'expérience. S'ils sont capables de recommander la stratégie d'exécution la plus rapide, ils continuent de préconiser la même stratégie même si elle ne fonctionne pas comme prévu.

En revanche, Db2 Machine Learning Optimizer intègre le retour sur la performance réelle de la requête pour suggérer des stratégies d'exécution qui donnent véritablement les meilleurs résultats sur votre infrastructure de données et affinent le chemin de requête à chaque exécution.

Avec un optimiseur Machine Learning, les DBA n'ont plus à surveiller la performance du système ni à tenter d'optimiser les requêtes. Ils peuvent se concentrer sur des activités à plus grande valeur ajoutée pour l'entreprise, comme mettre en place des applications d'IA, développer des stratégies d'utilisation des données et aider les utilisateurs professionnels de l'organisation à mieux exploiter les données disponibles.

Principe de fonctionnement

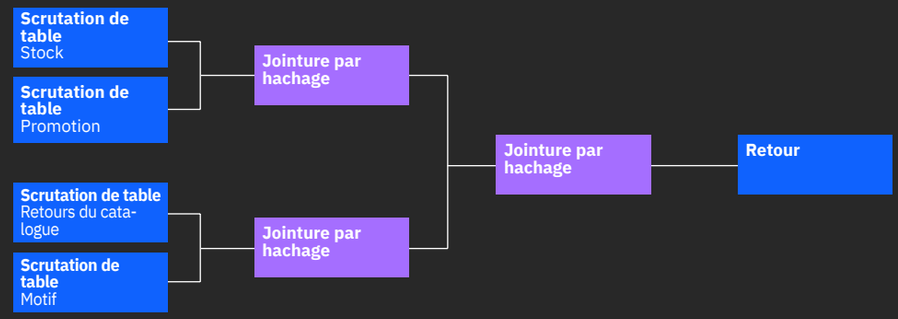
Db2 Machine Learning Optimizer reproduit les réseaux neuronaux pour optimiser les chemins de requête et traiter certaines d'entre elles 8 à 10 fois plus rapidement (tests internes d'IBM).

En action

L'optimisation des requêtes de ML

Un optimiseur de coûts traditionnel utilise la modélisation statistique et de ressources pour évaluer les stratégies d'exécution d'une requête donnée. Il renvoie la première option affichée sur la figure 2 ci-dessous : joindre deux paires de tables, puis joindre les résultats pour fournir un retour. Cependant, un algorithme de Machine Learning peut apprendre avec l'expérience et identifier une meilleure stratégie d'exécution : joindre une paire de tables, joindre une troisième table à ce résultat, puis joindre une quatrième à ce résultat pour fournir le même retour.

Sans le Machine Learning



Avec le Machine Learning

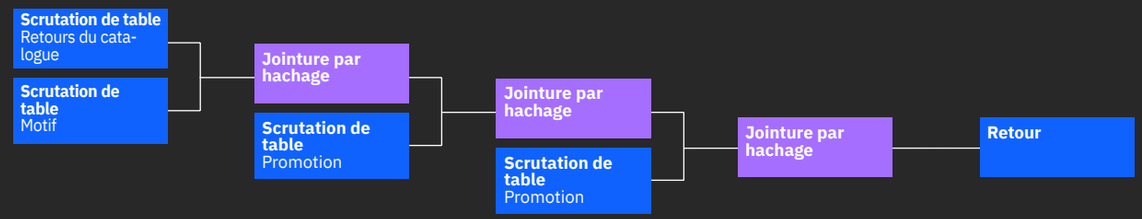


Figure 2. Exemple de stratégies d'exécution de requête, optimisées selon le coût (en haut) et par Machine Learning (en bas)

Ces changements de stratégie d'exécution peuvent être appliqués en temps réel, en réponse aux exécutions réelles des requêtes. Pour les applications urgentes comme la détection de fraude, cette amélioration en temps réel peut être critique.

Requête basée sur la confiance

Définition

Une fonctionnalité de pointe qui fournit des résultats de requête SQL sous la forme de probabilités (ou « meilleures correspondances ») plutôt qu'une simple réponse oui ou non.

Intérêt

La correspondance basée sur la confiance était disponible dans les paramètres de Machine Learning, mais cette fonctionnalité Db2 l'étend également aux expressions SQL. Pour un ingénieur SQL, cette fonctionnalité élargit l'éventail des tâches possibles sans faire appel à un « data scientist », ce qui rend les ingénieurs SQL encore plus précieux pour l'entreprise et allège la charge de travail, déjà énorme, qui pèse sur les « data scientists ».

Principe de fonctionnement

La requête basée sur la confiance ajoute des extensions de Machine Learning à SQL en mettant en œuvre des réseaux DFF (Deep Feed-Forward). Cette variante de Machine Learning non supervisé identifie les zones où la probabilité d'une correspondance est élevée. Le traitement des requêtes basé sur la confiance s'applique aux :

- Requêtes de similarité/dissimilarité
- Requêtes à raisonnement inductif (par exemple, clustering sémantique, analogie, mise à l'écart ou « odd-man-out »)
- Opérations de regroupement sémantique
- Anomalies comportementales (par exemple, détection de la fraude)
- Images, son, vidéo

En action

Requête basée sur la confiance

Une base de données de la police prend en compte de nombreuses variables : taille, poids, âge, couleur de la peau, caractéristiques distinctives, etc. Toutes ces données sont virtuellement incertaines en raison du manque de fiabilité des témoins. Dans une base de données traditionnelle, la recherche utilise une requête SQL très complexe pour prendre en compte cette incertitude en créant manuellement une « plage » de valeurs qui correspondra à la déclaration d'un témoin. Par exemple, un poids signalé de 88 kg peut correspondre à des suspects dont le poids varie entre 88 et 93 kg. Ces plages sont compliquées à manier et peuvent conduire à éliminer à tort des suspects si une valeur fournie par un témoin est trop éloignée de ladite plage. En revanche, un officier de police qui tente de faire correspondre les déclarations de témoin à un profil de suspect peut utiliser une requête basée sur la confiance pour générer une instruction SQL probabiliste qui identifie le sujet correspondant au mieux au profil donné par le témoin, même lorsque certains points de données sont hors de la plage d'origine des correspondances acceptables.

Conçue pour l’IA

Db2 est également conçu pour l’IA, ce qui permet à vos spécialistes des données de plus facilement mettre à niveau les charges de travail et les workflows pilotés par IA. Notamment :

- Exploration plus rapide des données avec des requêtes en langage naturel
- Utilisation de langages populaires et d’exemples de code pour créer des applications d’IA
- Modélisation de relations complexes à l’aide du graphe et de SQL
- Analyse des données de la Blockchain en mode natif

Exploration plus rapide des données avec des requêtes en langage naturel

Définition

Augmented Data Explorer est une interface web de type « Google », conçue pour explorer des données métier complexes. Elle permet aux utilisateurs d'effectuer des recherches en langage naturel et offre des API REST aux développeurs. Une question posée en langage naturel va générer des visualisations de données et des résumés qui utilisent toutes les sources de données disponibles.

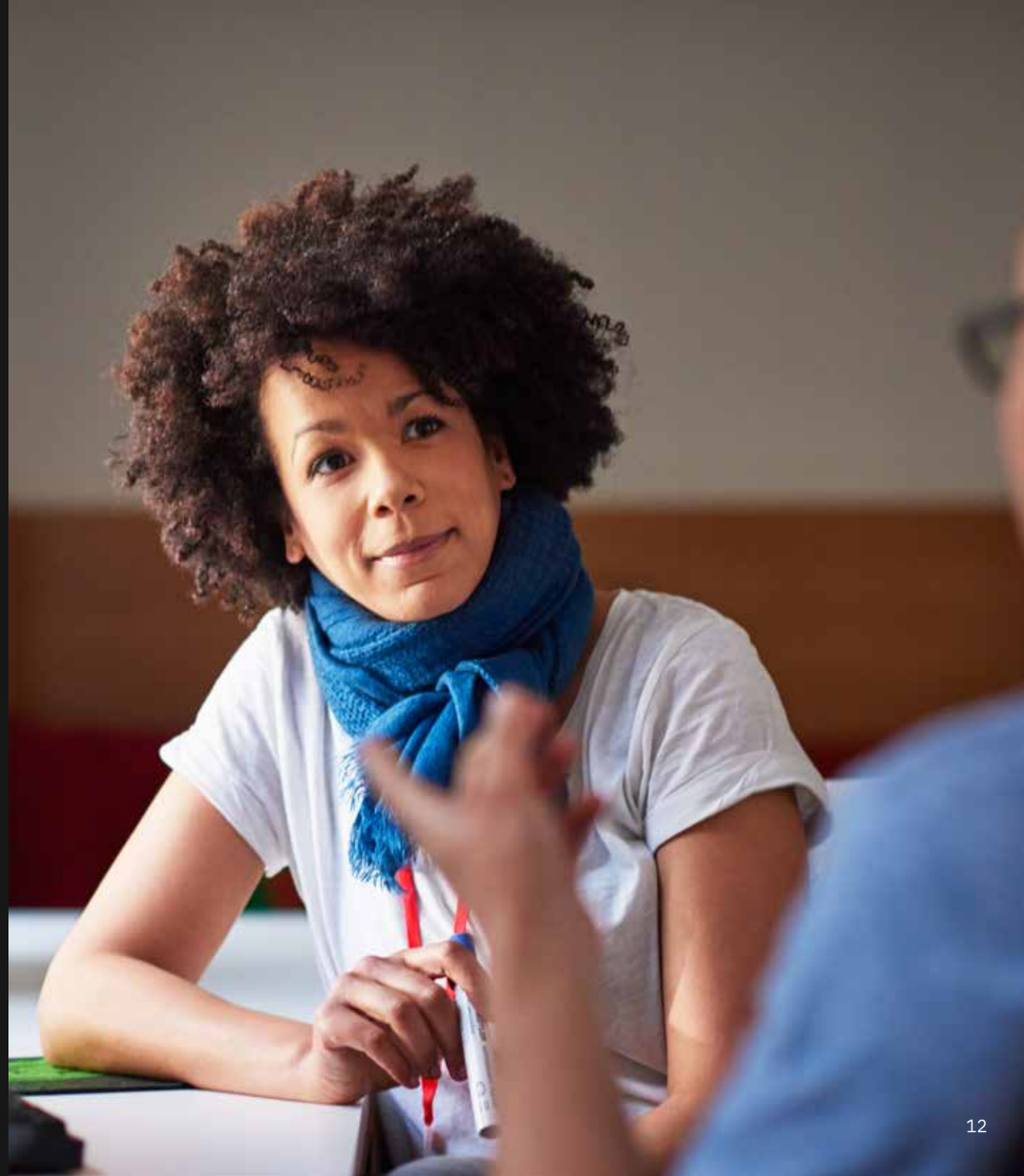
Intérêt

Alors que le paysage des données se complexifie toujours plus, de plus en plus d'utilisateurs sont confrontés à des ensembles de données au volume croissant. Les compétences nécessaires pour développer des requêtes SQL ou des applications d'analyse sont rares, mais les personnes qui veulent des réponses rapidement à partir de leurs données sont de plus en plus nombreuses.

Augmented Data Explorer est un outil d'exploration de données qui n'en est qu'à ses débuts. Il met les requêtes de données sophistiquées à la portée des utilisateurs non techniciens, qui n'ont plus qu'à entrer une requête en langage naturel comme « ventes en 2019 » ou « combien de fois pleut-il en octobre ? ». Cela va générer une instruction SQL qui interroge toutes les données pertinentes et renvoie le résultat souhaité sous la forme de langage naturel, d'une visualisation des données dynamique ou des deux.

Auparavant, les utilisateurs professionnels devaient suivre plusieurs processus manuels pour convertir des requêtes en langage naturel en une requête SQL, et faire appel à des experts pour corriger les éventuels problèmes. Augmented Data Explorer est conçu pour contourner ces deux obligations et trouver des solutions pour extraire plus de valeur des données disponibles.

Les « data scientists » et les utilisateurs plus expérimentés peuvent utiliser Augmented Data Explorer pour exécuter des requêtes initiales sur des ensembles de données inconnus ou volumineux, et savoir ce que ces derniers contiennent et où commencer les recherches. Augmented Data Explorer offre également des API REST pour que les développeurs puissent intégrer facilement ces fonctionnalités de recherche et de prévision en langage naturel dans leurs propres applications, au lieu de passer du temps à écrire le code de leur propre solution.



Principe de fonctionnement

Augmented Data Explorer utilise la technologie de ML pour parcourir les données, les indexer et en extraire des connaissances. Ses fonctionnalités d'AI se concentrent sur la prise en charge du langage naturel – comprendre le contexte, les synonymes et la syntaxe dans une grande variété de disciplines et de domaines – ainsi que sur l'anticipation des centres d'intérêt à partir des termes de recherche. Conteneurisé pour faciliter le déploiement et la gestion, l'outil dispose d'une interface web simple.

De plus, il intègre une fonctionnalité de graphe, qui indique en un clin d'œil les ensembles de données que le moteur a consultés et analysés. L'utilisateur peut facilement voir les données (schémas et tables) utilisées pour obtenir les réponses fournies par le système.

La valeur d'Augmented Data Explorer

Augmented Data Explorer est très précieux lorsque vos utilisateurs de données hésitent sur les questions à poser. Il aide ces utilisateurs à formuler leurs questions, à découvrir des relations au sein de ces données et à en extraire des connaissances exploitables. De plus, il permet à de plus en plus de personnes de relever ces défis métier. Dans des environnements métier en constante évolution, où les utilisateurs ont de nombreuses responsabilités, Augmented Data Explorer représente un outil de première ligne pour exploiter les volumes massifs de données disponibles actuellement et les valoriser davantage.

Prenons l'exemple du directeur des ventes d'un établissement commercial. Cette personne peut avoir différentes questions d'un jour à l'autre ou d'un mois à l'autre : répartition des ventes par région, analyse des taux de rétention des clients, identification des causes d'abandon de panier d'achat afin de savoir pourquoi des clients réguliers disparaissent, analyse de la relation entre météorologie et délais de livraison, etc.

Augmented Data Explorer est conçu pour identifier les nouvelles questions et y répondre rapidement, ouvrant ainsi les fonctionnalités d'IA à un panel bien plus large d'opérations métier.

Utilisation de langages populaires et d’exemples de code pour créer des applications d’IA

Définition

Db2 11.5 prend en charge en mode natif les bibliothèques et langages les plus courants : Python, JSON, GO, Ruby, PHP, Java, Node.js, Sequelize et les notebooks Jupyter. Les spécialistes des données peuvent créer des applications d’IA dans un langage qu’ils connaissent, puis les connecter de manière transparente aux données dans Db2.

Intérêt

Grâce à la prise en charge en mode natif des langages de Machine Learning les plus répandus, IBM crée un pont entre la gestion des données et le développement rapide d'applications avec un écosystème de données piloté par le ML et l’IA. Les développeurs et les « data scientists » peuvent conserver leurs données dans une base de données d’entreprise et conserver leurs bibliothèques et leurs langages de développement.

La prise en charge des bibliothèques et des langages en mode natif offre de nombreux avantages qui accélèrent et facilitent le développement d'applications d’IA :

- Les développeurs peuvent s'appuyer sur les bibliothèques et langages populaires, ainsi que sur certains exemples de code sectoriels sur GitHub, pour créer rapidement des applications répondant à leurs besoins.
- Les développeurs écrivent facilement des applications qui utilisent des données Db2 en mode natif, ce qui évite de personnaliser le code d’arrière-plan pour accéder aux données Db2, et supprime un point commun de défaillance ou de gestion pour leurs applications.
- Les nouvelles bases de données sont rapidement déployées et dimensionnées sur le Cloud en cas de besoin, puis connectées de manière transparente aux applications.
- Les développeurs ont accès aux environnements de développement de Machine Learning (IBM Watson® Studio) dans le même écosystème de bases de données.
- Les « data scientists » peuvent consacrer du temps à travailler sur les modèles de Machine Learning au lieu de résoudre des problèmes d’accès aux bases de données.
- Les entreprises peuvent recruter du personnel capable d’analyser les données sans avoir à acquérir de nouvelles compétences ou à apprendre de nouveaux langages simplement pour accéder d'abord aux données.

Principe de fonctionnement

Le processus de développement d'applications d'IA soulève plusieurs questions importantes :

– Où les données pertinentes sont-elles stockées ?

– Comment puis-je y accéder et les analyser ?

– Comment mettre en œuvre les pratiques de Machine Learning lors de la création d'une application ?

– Quels langages puis-je utiliser ?

L'écosystème IBM fournit des réponses prêtes à l'emploi à toutes ces questions. Les données peuvent être stockées dans Db2 ou Db2 Warehouse, Event Store ou Hadoop, dans le Cloud ou « on-premises », c'est-à-dire dans quasiment n'importe quelle combinaison de bases de données. Leur point commun, c'est leur adéquation aux besoins de l'entreprise, avec un support, une fiabilité et une performance maximum. La base de code Db2 commune et le moteur SQL commun sous-jacent permettent aux applications d'interroger les données de n'importe quel composant de l'écosystème, sans les modifier pour les exécuter sur différents éléments de cet écosystème. De plus, avec une série de nouveaux pilotes pour les principaux frameworks et langages de programmation Open Source, les développeurs peuvent facilement analyser et créer des modèles de Machine Learning dans les applications à l'aide de Db2.

Les spécialistes des données peuvent utiliser certaines des autres fonctionnalités décrites dans cet e-book, comme Augmented Data Explorer, pour comprendre rapidement ce que les données contiennent. Une fois stockées dans Db2, les données peuvent être connectées directement à l'environnement de développement de Machine Learning dans Watson Studio. Les bibliothèques sectorielles, les exemples de code et les modèles donnent aux développeurs une longueur d'avance dans le développement d'une application fonctionnelle. Enfin, un ensemble de pilotes natifs convertit les instructions SQL en un modèle propre au langage de ML (comme Python) en vue d'un développement ultérieur dans le langage choisi.

L'importance de la prise en charge des langages en mode natif

Le marché potentiel des applications d'IA est gigantesque. Jugez plutôt :

Distribution : moteurs de recommandation, contenus ciblés, promotions

Assurance : analyse de la fraude, traitement des applications

Logistique : gestion de parcs, maintenance préventive

Transports : conseil en temps réel d'itinéraire et de trajet selon le trafic, le temps, etc.

Une multitude de développeurs et de « data scientists » travaillent à développer des applications et des algorithmes intelligents, généralement en Python, Java ou un autre langage populaire. Les données qui alimentent ces applications résident dans une ou plusieurs bases de données. Certaines bases se connectent en mode natif à Python, mais en général, leur dimensionnement, leurs performances et leur fiabilité ne satisfont pas les entreprises. Pourtant, sans intégration native entre les langages de développement de ML et la base de données d'entreprise sous-jacente, il est difficile pour les développeurs de connecter l'application aux données nécessaires.

Grâce à l'intégration native entre Db2 et les langages de développement populaires, le développeur a accès aux données sans faire appel à un DBA ou à un spécialiste du SQL. Ce qui se traduit par une productivité accrue des développeurs et des « data scientists », ainsi que par des applications plus robustes et développées plus rapidement.

L’IA permet auxutilisateurs d’exploiter les données à un niveau inédit.

Modélisationderelationscomplexes à l’aide du graphe et de SQL

Définition

Db2 intègre profondément la fonctionnalité de graphe dans les données relationnelles et SQL, afin que les applications s’exécutent directement à partir des données relationnelles et que le moteur SQL de Db2 puisse interroger directement les données du graphe.

Intérêt

Le graphe est un outil impressionnant qui fournit des informations importantes, mais qui était incompatible avec les systèmes transactionnels et OLTP. Pour résoudre cette incompatibilité, les organisations gardaient leur infrastructure relationnelle, mais extrayaient certaines données avant de les intégrer dans une base de données de graphe et de créer des applications de graphe dessus. Ce type de duplication, de retard et de traitement n’est plus acceptable. Les organisations ont besoin de connaissances pilotées par le graphe dans leurs données relationnelles, et cela en quelques secondes.

Dans de nombreux secteurs, les utilisateurs veulent ajouter des fonctionnalités de graphe dans leur kit d’outils d'analyse, mais sans perturber leur infrastructure de données ni ajouter une nouvelle base de données uniquement pour prendre en charge le graphe. Avec Db2, c’est possible.

Principe de fonctionnement

Les données de graphe sont stockées dans des tables du framework relationnel de Db2 ou dans Db2 Event Store. Autrement dit, le moteur SQL peut interroger les données du graphe directement, et les applications de graphe peuvent interroger les données directement dans les tables relationnelles.
Cette architecture prend également en charge les langages d’interrogation de graphe Open Source, comme Gremlin et Tinkerpop.

L’intégration étroite entre les données relationnelles et les fonctionnalités de graphe dans Db2 offre de nombreux avantages de poids :

- Le graphe peut s’exécuter directement sur les données relationnelles existantes, ajoutant un autre niveau d'analyse en plus de l'analyse SQL.
- L'analyse SQL peut s’exécuter directement sur les données du graphe stockées dans les tables relationnelles.
- Les applications de graphe peuvent utiliser une API personnalisée ou Spark pour se connecter à Db2.
- Les transactions ACID mettent à jour le graphe en temps réel sans perturber les applications relationnelles existantes, ce qui permet d’utiliser les graphes à la fois pour l'analyse et les transactions en temps réel.
- Certaines solutions sectorielles exploitent les graphes de manière intensive, comme la santé et la finance.

L’importance de la prise en charge en charge du graphe en mode natif

Souvent, les analyses fondées sur les graphes permettent de gagner en efficacité et d’exécuter des nouveaux workflows sur des données relationnelles. Par exemple, un détaillant pourrait utiliser des relations de graphe en temps réel pour identifier le dénominateur commun des retours d’un article donné sur plusieurs sites. Ensuite, cette information peut être intégrée immédiatement la base de données transactionnelle du détaillant et entraîner le retrait de cet article des rayons dans la journée.

Dans le secteur de l’assurance, un enquêteur pourrait utiliser le graphe pour fusionner plusieurs ensembles de données : nombre et taille des réclamations, relations entre les demandeurs, identités des fournisseurs de services, etc. Les relations formalisées dans le graphe peuvent mettre en évidence une probabilité de fraude.

Analyse des données de la Blockchain en mode natif

Définition

L'API native Blockchain Connector dans Db2 offre une transparence et une capacité d'analyse inédites sur les données ultra-compressées des livres de la Blockchain.

Intérêt

La technologie de la Blockchain s'impose rapidement dans de nombreux secteurs, mais elle génère des goulets d'étranglement au niveau de l'analyse. Des volumes massifs d'informations potentiellement utiles sont stockés sous forme compressée dans chaque livre de la Blockchain, mais il n'existait aucun moyen simple de consulter et d'analyser ces informations. Db2 Blockchain Connector modifie cette donne. Compacts, transparents, contrôlables et immuables, ces livres de la Blockchain sont maintenant analysables, au même titre que n'importe quelle autre source de données dans l'entreprise.
Les entreprises profitent de tous les avantages du moteur Db2 – performance, flexibilité, évolutivité et sécurité –, sans avoir à développer des solutions ad hoc de reporting pour la seule Blockchain.

Db2 Blockchain Connector va même au-delà, avec la connectivité et l'interopérabilité natives des autres types de données disponibles dans Db2. Les entreprises peuvent facilement intégrer des données d'autres bases pour compléter le contexte et ouvrir de nouvelles possibilités pour leurs données de la Blockchain. Enfin, ce connecteur permet d'exploiter les données de la Blockchain dans les applications d'IA, une utilisation qui aurait été compliquée auparavant en raison du cloisonnement des données de la Blockchain. Désormais, les développeurs d'IA peuvent intégrer facilement les ensembles de données de la Blockchain, soit comme source de données primaire pour leurs applications, soit pour fournir des détails encore plus précis.

Principe de fonctionnement

Db2 Blockchain Connector balaie les données transactionnelles compressées et stockées dans le livre de la Blockchain, puis les présente sous la forme d'une table relationnelle dans Db2.
La robuste stratégie de mise en cache utilise les fonctionnalités Db2 pour créer une table en cache des données de la Blockchain pour améliorer les performances de la requête, tout en permettant aux utilisateurs d'obtenir les données les plus récentes en cas de besoin.

L'importance de la prise en charge de la Blockchain en mode natif

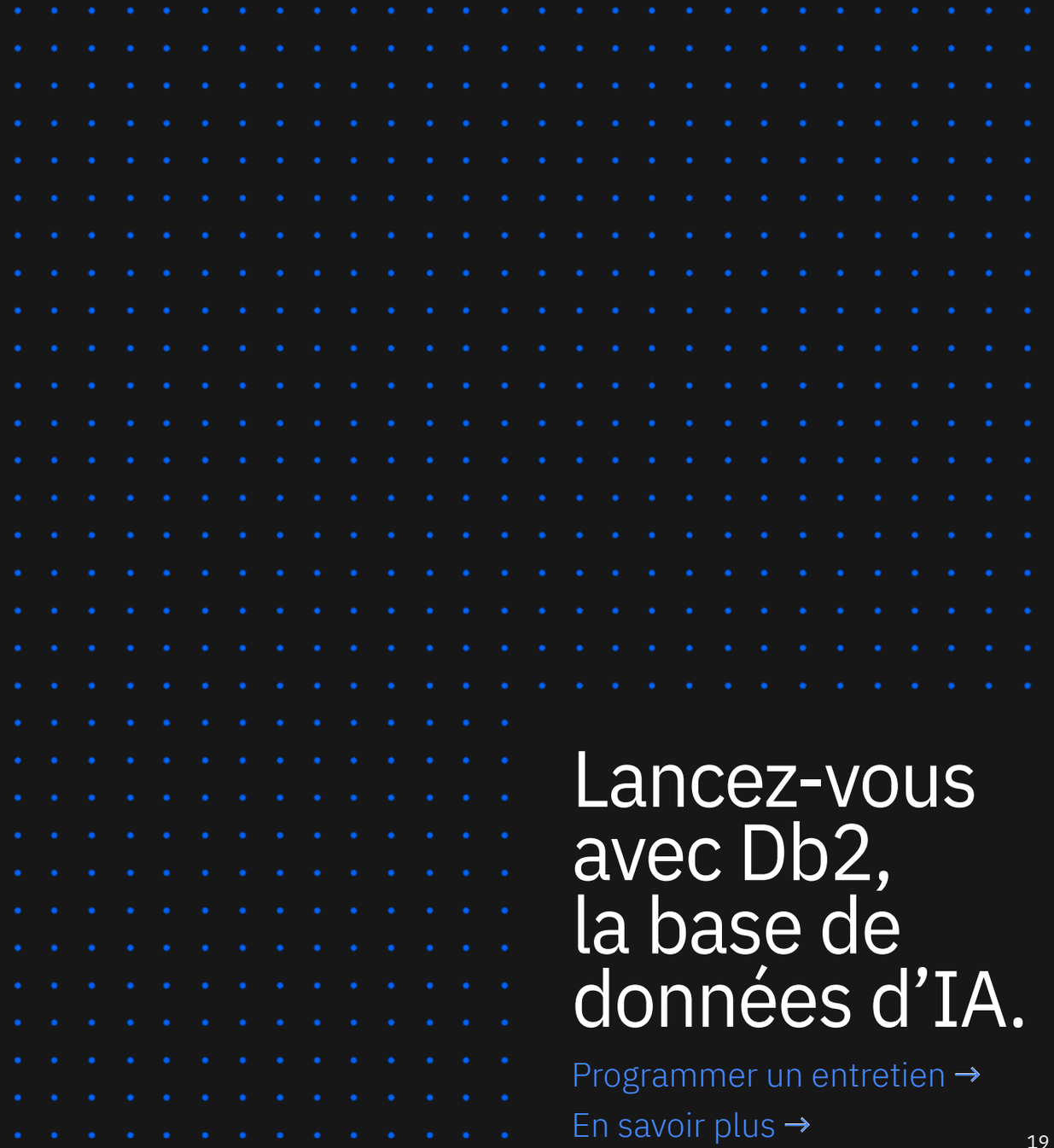
La Blockchain est déjà utilisée dans des secteurs comme l'assurance, la finance, la logistique et la santé pour stocker les transactions immuables, mais les acteurs de ces secteurs veulent également effectuer des analyses pour mieux comprendre leurs données. Par exemple, les compagnies d'assurances veulent savoir combien de réclamations ont été déposées par client dans une région donnée ; les sociétés de transport veulent suivre leurs conteneurs et leurs véhicules ; et les établissements de santé veulent comprendre les dossiers de santé de leurs patients.

Au lieu de créer une application d'arrière-plan personnalisée pour extraire ces données, ces organisations peuvent simplement connecter Db2 à leur base de données de la Blockchain et analyser ce contenu comme elles le feraient pour n'importe quel autre. Grâce à la connectivité et à la flexibilité inhérentes de Db2, elles peuvent corréler des données contextuelles supplémentaires aux données transactionnelles de la Blockchain (par exemple, la météorologie pendant une livraison particulière, les collaborateurs d'un demandeur qui auraient pu avoir déposé des réclamations similaires) et obtenir une vue plus riche et plus complète de l'impact de chaque transaction sur l'entreprise. Les applications d'IA peuvent désormais s'exécuter sur les données de la Blockchain.

La place de Db2 dans une organisation pilotée par l'IA

Que vous déployiez des fonctionnalités Db2 dans le cadre d'un déploiement « on-premises » traditionnel, dans une instance d'entrepôt ou de base de données relationnelle dans le Cloud, sur la plateforme Hybrid Data Management Platform disponible sur abonnement, sur la nouvelle solution IBM Cloud Pak™ for Data ou sur un autre modèle de déploiement, la technologie sous-jacente est passionnante et les avantages métier sont considérables.

Db2 contribue à faire de l'IA une réalité pour les utilisateurs dans les entreprises de toute taille. La proposition de valeur est claire. Créez, mettez en place et gérez une IA d'entreprise sur n'importe quelle infrastructure : « on-premises » ou dans le Cloud. Exploitez toutes vos sources de données, où qu'elles soient physiquement. Tirez parti des fonctionnalités d'IA intégrées dans la base de données pour obtenir des connaissances prédictives et proactives permettant de prendre des décisions fondées sur les données. Et avec Db2 comme base, les entreprises non seulement créent et connectent les applications d'IA plus efficacement, mais elles en extraient des informations exploitables rapidement et rendent l'IA accessible à un public plus large que jamais.



Lancez-vous avec Db2, la base de données d'IA.

Programmer un entretien →

En savoir plus →

IBM France
17 Avenue de l'Europe
92275 Bois Colombes Cedex

Produit aux États-Unis, janvier 2020.