



LIVRE BLANC

IBM DB2 pour SAP : plate-forme de base de données analytiques et transactionnelles sans compromis

Sponsorisé par : IBM

Carl W. Olofson
Dan Vesset
Avril 2014

Matthew Eastwood

RÉSUMÉ

Ce livre blanc est rédigé pour les clients SAP évaluant leurs choix d'infrastructure, abordant l'évolution de la technologie des bases de données et les options disponibles. Il n'est pas facile de mettre en avant un choix binaire car les charges de travail de SAP chevauchent l'analyse en temps réel et le traitement des transactions extrêmes, et les choix d'infrastructure peuvent à présent être vastes, étant données les avancées de la technologie en matière de traitement en mémoire et de vitesses de traitement plus rapides.

IBM propose DB2 avec BLU Acceleration pour une performance extrême sans limite de l'analyse et DB2 PureScale pour des capacités d'évolutivité complètes et flexibles de SAP ERP. Selon IBM, ces configurations DB2 exécutées sur Power Systems forment la solution complète pour tous les clients SAP. En prenant en compte toutes les options du marché, ainsi que leur maturité, IDC recommanderait aux utilisateurs SAP d'envisager d'adopter IBM DB2 comme plate-forme de données pour toute la gamme d'applications SAP, aussi bien transactionnelles qu'analytiques.

AVIS D'IDC

Un grand nombre de sociétés dans un grand nombre d'industries dans le monde entier utilisent des applications SAP comme principal support pour leurs processus stratégiques. Les recherches d'IDC indiquent que ces sociétés font face à une pression continue pour prévoir et fournir une plate-forme informatique qui s'adapte à la demande de traitement que ces charges de travail SAP exercent. C'est-à-dire que les responsables de processus opérationnels internes et les clients, partenaires et fournisseurs, se fiant tous à ces applications, exigent en permanence une plus grande réactivité et un accès plus important aux données sensibles. Les meilleures extensions mobiles de SAP dans l'industrie, telles que MEM ERP, ont contribué en outre au volume, à la variabilité et aux caractéristiques (par ex., changement des rapports lecture/mise à jour) de ces charges de travail. Ensuite, la société a besoin d'une synthèse encore plus rapide et plus large des données émergentes issues d'applications SAP et d'autres sources (CRM, consultation sur Internet, etc.) pour l'analyse. La partie la plus lourde de ces deux exigences se trouve directement au cœur du système : le composant central SAP ERP (SAP ECC) et Business Information Warehouse (BW)] et la base de données qui les exploite. Pour avancer et réussir, les services TI doivent analyser ces principaux éléments de base de données essentiels de la manière la plus efficace, la plus rentable et la plus rationalisée possible. En bref,

ces demandes transactionnelles et analytiques requièrent des bases de données faciles à administrer et plus puissantes que jamais.

Afin de garantir que la base de données d'une application SAP gère ces demandes transactionnelles et analytiques variables et croissantes, sans surcharger le personnel du centre de données ou les administrateurs de la base de données (DBA), les sociétés doivent mettre en oeuvre une plate-forme qui peut :

- Assurer une gestion des données qui est améliorée grâce à un système de compression économisant de l'espace, simplifiant le fonctionnement et accélérant le traitement par rapport à l'optimisation classique des disques.
- Organiser les données analytiques en colonnes.
- Prendre en charge l'évolutivité dynamique des données transactionnelles avec une disponibilité importante grâce au groupage.
- S'adapter au plan de gestion global du centre de données, appartenant idéalement à la même gamme de produits que celle utilisée ailleurs pour la gestion des bases de données stratégiques.

IBM DB2, avec sa technologie d'optimisation pour les transactions et l'analyse, se présente comme principal candidat à cet égard, et une valeur supplémentaire peut être obtenue en l'exploitant sur IBM Power Systems.

Ce livre blanc aborde les exigences des bases de données pour les applications SAP car le rythme accéléré de l'activité pousse les entreprises à chercher de l'aide pour le traitement des transactions extrêmes et l'analyse en temps réel. Il souligne les principaux développements au niveau de la technologie des systèmes de gestion des bases de données (DBMS) qui abordent ces exigences. Il montre comment DB2, avec PureScale, permet un traitement des transactions extrêmes et comment DB2 avec BLU Acceleration offre la puissance pour une analyse en temps réel. Il montre également comment DB2 sur Power Systems aborde globalement les besoins des clients SAP sans les forcer à se compromettre dans l'organisation et le fonctionnement du centre de données.

PRÉSENTATION DE LA SITUATION

Tout le monde connaît bien maintenant les problèmes impliqués par le rythme accéléré du business : l'exigence de disponibilité de la société 24 h sur 24 et 7 jours sur 7, la vitesse du traitement des transactions due à la numérisation de pratiquement tout, et l'hyper-compétitivité des entreprises due à la disponibilité du suivi des clients et des données relatives aux opinions issus des appareils mobiles, du suivi des clics Internet et de diverses sources de données volumineuses (y compris, tout en ne s'y limitant pas, les médias sociaux). La pression pour prendre des décisions plus éclairées sur le moment, pour conclure des affaires plus rapidement et pour éliminer les périodes d'indisponibilité a entraîné deux principales zones d'amélioration fonctionnelle sur les applications commerciales du passé. Il s'agit du traitement des transactions extrêmes et de l'analyse en temps réel.

Technologie de base de données requise pour l'entreprise en temps réel

Traitement des transactions extrêmes

Auparavant, les sociétés rythmaient leur activité, en partie, selon le rythme auquel leurs systèmes informatiques pouvaient traiter les transactions. Ceci n'est plus acceptable. À la place, les systèmes informatiques doivent traiter les transactions à un rythme qui répond au rythme de l'activité. Ainsi, la base de données, associée à une

architecture conçue pour la performance, doit permettre à l'entreprise de travailler au rythme pratique le plus rapide du point de vue des clients, partenaires et fournisseurs.

Pour répondre à cette exigence, la base de données doit exécuter même les transactions complexes à un rythme garantissant qu'il ne ralentit pas le processus commercial. Un tel rythme peut nécessiter des vitesses de transaction allant d'une durée inférieure à la seconde à quelques millisecondes. Pour la plupart des DBMS, cela a impliqué un remaniement radical de la manière dont les données transactionnelles sont gérées afin de garantir l'évolutivité avec une disponibilité et un débit maximaux.

Analyse en temps réel

Outre la nécessité de transactions rapides, les sociétés requièrent également des transactions plus éclairées. L'accélération des transactions peut entraîner la multiplication des erreurs sauf si davantage d'informations sont introduites dans le processus. L'exécution de requêtes complexes pertinentes pendant une transaction de la société peut activer des opérations commerciales plus efficaces et davantage de ventes. Par exemple, si un ingénieur commercial peut trouver des schémas de vente (grâce à une requête complexe très rapide) selon le comportement des clients correspondant au profil du client qui se trouve là, il peut prendre une décision sur le moment conduisant à une meilleure interaction de vente. Le recueil actuel de données peut permettre d'obtenir des informations commerciales, conduisant à des interactions contextuelles plus riches avec les clients, partenaires et fournisseurs, ce qui entraîne une efficacité opérationnelle supérieure et une meilleure exploitation de l'opportunité commerciale.

La capacité à fournir des informations commerciales opportunes aussi rapidement que les décisions tactiques doivent être prises, et avec les données les plus récentes, s'appelle « analyse en temps réel » car les analyses sont si sensibles au temps qu'elles doivent être fournies dans l'intervalle de temps valable de l'activité commerciale réelle. Elles requièrent des technologies DBMS spécifiques comme nous le verrons dans les sections suivantes.

Changements au niveau du paysage technologique

Ces exigences commerciales, associées aux changements de technologie d'un point de vue de la capacité et du coût, entraînent le développement et l'offre d'un DBMS significativement transformé, qui est destiné à activer le traitement des transactions extrêmes et l'analyse en temps réel. Pour comprendre comment cela est possible, il faut prendre en compte les récents développements de l'évolution de la technologie informatique.

Clustering en vue de l'évolutivité et du basculement

La capacité d'une base de données à évoluer avec flexibilité, selon le volume de données traitées, est essentielle pour permettre aux entreprises de garder le rythme de l'activité. L'approche essentielle de l'évolutivité des bases de données transactionnelles dont le coût est gérable implique des clusters. L'architecture en cluster des bases de données existe depuis assez longtemps, mais les techniques requises pour permettre le traitement des transactions en cluster tout en gérant efficacement les verrous afin d'éviter les retards ne sont disponibles que dans quelques produits DBMS. Cependant, une telle architecture est essentielle à la réussite d'une base de données transactionnelles dynamiquement évolutive. Un cluster de stockage partagé qui gère des verrous et partage une mémoire tampon efficacement entre des nœuds sans un seul point de défaillance est essentiel pour fournir un serveur de base de données transactionnelles pouvant évoluer et garantir une disponibilité continue.

Un certain nombre de fournisseurs de DBMS majeurs prétendent fournir cette technologie. En évaluant ces revendications, il est important de prendre en compte comment elles sont obtenues et la mesure dans laquelle les fournisseurs les concrétisent. Les facteurs clés à rechercher comprennent :

- Tous les nœuds actifs et pouvant contribuer au problème en question tout en agissant également comme nœuds de basculement si on l'exige.
- Aucun congestionnement ni aucun conflit entre les nœuds pouvant ralentir ou arrêter l'intégralité du groupe.
- La simplicité d'administration y compris la composition DDL, l'application SQL et l'ajout ou la suppression de nœuds.
- La capacité à ajouter des ressources supplémentaires pour gérer les « pics » de demande et à les réduire ensuite.

Grands espaces mémoire et processeurs rapides

Auparavant, des quantités relativement limitées de mémoire vive à accès direct (DRAM) et la vitesse comparativement plus lente des processeurs limitaient la performance DBMS. Au cours des 10 années passées, les processeurs ont significativement augmenté leur puissance et peuvent à présent utiliser 64 bits pour la mémoire. Cela signifie qu'ils peuvent gérer des espaces mémoire beaucoup plus importants. En outre, chaque processeur comporte plusieurs cœurs, garantissant que chaque processeur peut exécuter autant de processus purement parallèles qu'ils ont de cœurs. En même temps, le coût des processeurs et de la DRAM a baissé, les principales architectures de bus se sont améliorées pour offrir de plus grandes vitesses d'accès pour faire face aux grandes quantités de mémoire, et de meilleurs jeux d'instructions des processeurs permettent une utilisation plus efficace du cache des processeurs, réduisant ainsi le nombre d'accès à la DRAM. Tout cela a augmenté de façon spectaculaire la taille et la vitesse du traitement en mémoire.

Passer d'une technologie de base de données optimisée sur disque à une technologie de base de données optimisée en mémoire

En raison des limites historiques inhérentes à la mémoire, au processeur et à l'espace adresse au cours des années précédentes, les DBMS ont concentré la gestion des données sur des disques afin d'optimiser leurs opérations. Ainsi, la technologie DBMS traditionnelle exerce une grande quantité de son activité de traitement à gérer des données sur disque, maintenir des tampons en mémoire et à adresser les données selon l'accès au disque. Il est à présent pratique de gérer des données en mémoire plutôt que sur disque. L'ancienne approche de la gestion des données est une gestion de base de données « optimisée sur disque », tandis que la nouvelle approche est « optimisée en mémoire ». L'optimisation en mémoire n'implique pas qu'une mise en cache ou en tampon intelligemment gérée, mais un changement fondamental de l'architecture de la base de données, de la couche de stockage vers le haut et optimisée pour des opérations en mémoire. Les bénéfices de la gestion de base de données optimisée en mémoire comprennent une administration plus simple et des opérations de base de données beaucoup plus rapides, tout particulièrement le traitement des requêtes.

Organisation des données basées en mémoire

La gestion des bases de données en mémoire permet au noyau de la base de données de réaliser des opérations de base de données basiques avec moins d'instructions, souvent moins d'un dixième du nombre requis pour une base de données optimisée sur disque. Prenons l'exemple d'un problème hypothétique simple de récupération de données. Pour récupérer des données, un système optimisé sur disque doit conserver des clés de la base de

données permettant au DBMS de déterminer l'endroit des pages des données ; déterminer si les données sont dans un tampon, et si ce n'est pas le cas, identifier où la page se trouve sur le disque ; récupérer la page et déloger une autre page pour lui faire de la place ; puis, traverser la page pour trouver le dossier associé. Au contraire, le DBMS optimisé en mémoire récupère les données en convertissant simplement la clé de la base de données (adresse physique réelle) en un pointeur en mémoire et va directement à l'endroit où se trouvent les données en mémoire.

Cela signifie que les requêtes complexes sont exécutées en un dixième du temps requis par les systèmes traditionnels optimisés sur disque, et que chaque nœud de serveur peut gérer six fois la charge que des systèmes traditionnels optimisés sur disque pourraient gérer. Les organisations mettant en œuvre cette technologie économisent car elles accélèrent les transactions qui ont des requêtes intégrées, augmentent la capacité de requête et réduisent l'empreinte des infrastructures, permettant à la société de faire plus d'affaires et de prendre des décisions plus éclairées et plus rapides.

Outre la meilleure efficacité du noyau DBMS, le DBMS optimisé en mémoire réduit le travail du personnel informatique impliqué dans la gestion de la base de données. Avec un DBMS optimisé sur disque, l'administrateur de la base de données doit développer une stratégie de gestion du stockage qui implique habituellement des tâches telles que l'attribution minutieuse de tablespaces à des fichiers sur des volumes, la distribution de données de table dans des volumes à l'aide de partitions hash key, en utilisant le striping et le mirroring des données pour une plus grande disponibilité, et en maintenant des indexes globaux et de partition pour une récupération plus rapide des données aléatoires. En outre, il existe un besoin régulier de réaliser des sauvegardes des points de contrôle, décharger et recharger des opérations lorsque les définitions de table changent, défragmenter la base de données et indexer les reconstructions. Une base de données optimisée en mémoire élimine toutes ces tâches. Une sauvegarde des points de contrôle simple et très compressée (ou sauvegarde d'instantané) sur un disque est généralement la seule tâche opérationnelle, et cela peut souvent être réalisé sans mise au repos de la base de données. Les seules autres écritures requises sur le disque sont en général des écritures faites sur un journal ou des journaux de changement, utilisées pour le retour en arrière, le recouvrement et la récupération.

Pour une base de données analytiques, une architecture optimisée en mémoire offre des avantages considérables. Cependant, pour une vaste base de données transactionnelles, une approche optimisée sur disque est souvent préférée car elle peut offrir une meilleure performance de transaction globale, une meilleure fiabilité et une plus grande disponibilité que les autres options. Il est donc nécessaire de choisir le bon outil pour le travail requis.

Transactions *versus* analyse de données

Les transactions requièrent généralement la récupération et l'insertion de toutes les lignes des tables, parfois dans le contexte de requêtes assez simples. De même, les transactions requièrent généralement l'insertion ou la mise à jour d'un petit nombre de tables, mais la requête d'un grand nombre de tables. Les analyses de données, d'autre part, impliquent des requêtes complexes impliquant des colonnes sélectionnées, appliquant parfois des critères de sélection mathématiques sur plusieurs tables, et imbriquées sur plusieurs niveaux de profondeur. En raison des différentes exigences de traitement des transactions et de traitement analytique, le DBMS doit organiser les données différemment pour ces deux classes de traitements. Dans le cas des transactions, les données sont généralement conservées sous forme de lignes. Pour le traitement analytique, une organisation en colonnes est préférée, et dans une base de données optimisée en mémoire, ce format peut être davantage optimisé en utilisant la compression et le positionnement minutieux des données à accès séquentiel sur des limites spécifiques d'adresse mémoire pour des avantages de performance supplémentaires.

Technologie d'optimisation des transactions

La gestion des tables en lignes avec optimisation de la mémoire dans le serveur de la base de données est la meilleure approche pour l'optimisation des transactions. Comme la plupart des applications de base de données transactionnelles ont tendance à ne pas faire référence aux mêmes données une fois écrites au même stade de traitement, la gestion de ces données sur disque constitue l'approche la plus fréquente pour maintenir les données plutôt que de les conserver en mémoire. Pour obtenir une évolutivité flexible, afin qu'au fur et à mesure que le volume des transactions augmente le serveur puisse être étendu pour faire face à la demande, une technique de clustering de disques partagés est nécessaire.

Technologies d'accélération des analyses de données

Pour les analyses de données, il existe un grand nombre de techniques pour mieux optimiser la récupération des données dans une base de données optimisée en mémoire.

Compression et organisation en colonnes

Comme nous l'avons précédemment mentionné, une organisation des données en colonnes est optimale pour la plupart des traitements analytiques. Il existe deux raisons à cela : la plupart des requêtes analytiques impliquent des critères faisant référence uniquement à un petit sous-ensemble des colonnes pour les tables impliquées dans leurs opérations, et en observant une colonne de données organisées ensembles, le DBMS peut profiter des schémas d'organisation et de compression réduisant le temps de requête et éliminant la nécessité d'index.

Traitement des vecteurs

En général, le traitement des vecteurs est une technique de programmation impliquant la configuration des vecteurs (tableaux unidimensionnels ou listes en série de valeurs contiguës) et permettant aux instructions uniques de fonctionner sur ces vecteurs à l'aide d'une approche appelée SIMD (instruction unique, données multiples). Grâce au SIMD, le travail peut être étendu sur plusieurs threads et cœurs de processeur pour accélérer grandement la performance, ce qui permet aux décideurs d'obtenir les réponses dont ils ont besoin rapidement. Plus il y a de cœurs et de threads par cœur, meilleure est la performance. Par exemple, la technologie sur processeur IBM POWER8 propose huit threads par cœur, deux fois plus que les autres offres. Ce plus grand nombre de threads doit pouvoir fournir une performance encore meilleure.

Associée à des colonnes, cela implique d'organiser les données en colonnes dans la mémoire, ainsi lorsque les processeurs agissent sur les données, elles sont chargées dans le cache des processeurs de telle manière que chaque processus puisse agir sur le nombre maximal de valeurs avant de retourner à la DRAM. Cette technique peut diviser le temps de recherche par 10 ou plus. Associée à une compression optimale utilisant une stratégie garantissant que les données n'ont pas à être décompressées avant de réaliser des évaluations comparatives, le gain de performance peut être beaucoup plus important.

L'illusion de la base de données unique

Une idée véhiculée par certains fournisseurs consiste à dire qu'il serait idéal de conserver les données transactionnelles et analytiques dans la même base de données. À première vue, cela semble faire sens : cela semble plus simple et plus facile à gérer. Cependant, en regardant de plus près, nous voyons que la vision d'une seule base de données transactionnelles et analytiques n'est que pure illusion.

Pour commencer, comme nous l'avons déjà vu, les technologies impliquées dans l'optimisation des transactions et des requêtes analytiques complexes sont très différentes. Nous avons vu que des clusters de nœuds qui

partagent des ressources et *gèrent les données en lignes* sont les plus judicieux pour le traitement des transactions, tandis que le fait de conserver *les données dans des colonnes compressées de manière optimale* offre de grands bénéfices pour l'analyse de données. Cependant, les raisons justifiant pourquoi gérer les données analytiques et transactionnelles séparément sont plus simples et plus fondamentales. Elles comprennent :

- Les données nécessaires pour exécuter une transaction et les données nécessaires pour une analyse ne sont pas les mêmes, même pour le même domaine. Les données transactionnelles comprennent des détails non importants pour l'analyse mais essentiels pour les transactions, tandis que les analyses ont besoin de données contextuelles qui sont inutiles au traitement des transactions. Le fait de les associer entraîne un schéma confus et très complexe.
- Les analystes peaufinent en permanence les données qu'ils analysent et ont besoin de liberté pour modifier le schéma sans craindre que ces changements puissent avoir un impact négatif sur le traitement des transactions.
- Pour les utilisateurs SAP, il n'existe vraiment pas de chevauchement, au moins dans un premier temps, entre les données gérées par SAP Business Suite et les données gérées par SAP Business Warehouse ou les outils analytiques SAP. Les données transactionnelles utiles doivent encore être converties en une forme qui les rend exploitables pour les requêtes analytiques et les rapports (et visualisations) opérationnels.

Il existe également des problèmes opérationnels. Même si une base de données peut conserver les données principalement transactionnelles en lignes et les données principalement analytiques en colonnes tout en gérant les deux avec le même schéma pour supporter des requêtes mixtes, il existe des compromis évidents de performance et d'efficacité qui découlent d'une telle approche, ainsi que certaines interdépendances opérationnelles pouvant se développer entre les données principalement transactionnelles et les données principalement analytiques : par exemple, des clés étrangères pourraient empêcher les optimisations conçues pour rendre chaque mode d'opération efficace.

Ceux qui promeuvent l'approche de la base de données unique disent que l'approche ETL (extraction, transformation et chargement) par lots pour déplacer les données du système transactionnel au système analytique entraîne un temps d'attente beaucoup trop long qui empêche les analyses en temps réel. Ceci est vrai, mais l'ETL par lots n'est pas la seule option. La transmission des données en mouvement basé sur une technologie telle que le Change Data Capture (CDC), peut garantir que la base de données analytiques est suffisamment actuelle pour les analyses requises.

CDC versus ETL

Il existe également une différence entre, d'une part les données analytiques utilisées pour les analyses en temps réel qui sont liées aux transactions et décisions « au point d'impact » et, d'autre part, les données analytiques qui sont utilisées pour l'analyse d'une période de temps ou d'une autre dimension clé. Cette dernière forme d'analyse est essentielle à la prise de décisions stratégiques sur le plus long terme. Tandis que le premier cas peut être couvert par un data store analytiques grande vitesse, le dernier cas peut nécessiter un entrepôt de données. Le premier peut exiger un mouvement des données dynamiques sur CDC de la base de données transactionnelles aux données analytiques, tandis que le dernier exige une opération ETL programmée impliquant des données issues de plusieurs sources pour le chargement dans un entrepôt de données.

Ceci dit, il est vrai que la gestion de toutes les données utiles sous une seule gamme de produits DBMS est très sensée, tout particulièrement si le fournisseur peut offrir un moyen de gérer les données dans les deux classes de bases de données de manière cohérente.

L'argument pour IBM DB2 pour SAP

En prenant en compte les exigences de la base de données pour les versions successives des applications SAP Business Suite, SAP Business Warehouse et SAP BusinessObjects, l'utilisateur dispose d'un certain nombre d'options possibles qui sont toutes raisonnables. La difficulté est de choisir la bonne option pour vos besoins. DB2 pour Linux, Unix ou Windows (LUW) est l'une de ces options et comprend plusieurs différentiateurs qui doivent être pris en compte.

IBM DB2 PureScale pour SAP Business Suite

IBM DB2 PureScale est conçu pour offrir le type d'évolutivité requise par un système d'application ERP complet tel que SAP Business Suite. Il fonctionne sur un cluster de stockage partagé permettant aux utilisateurs d'ajuster le nombre de nœuds pour s'adapter aux demandes de traitement presque sans aucun effort. Ses principales caractéristiques distinctives incluent :

- Une évolutivité qui n'a aucune limitation inhérente, permettant l'ajout de nœuds avec quelques commandes simples.
- Une gestion centralisée des verrous qui évite le problème que certains clusters de stockage partagé rencontrent avec des collisions de gestion des verrous pouvant immobiliser l'intégralité du cluster.
- La gestion centralisée des tampons qui permet à tous les nœuds utiles d'avoir des données actuelles en mémoire, grâce à une technique qui rafraîchit leurs tampons sans affecter leurs opérations ou leur performance : il s'agit d'une technique appelée accès direct en mémoire distant (RDMA), qui permet la mise à jour de la mémoire d'un système distant sans causer d'interruption.
- L'approche DB2 de gestion des données partagées qui écarte la nécessité de tout type de partitionnement pour obtenir une meilleure performance, contrairement à certaines autres technologies.

Ces avantages sont en partie le résultat de décennies d'expériences d'IBM dans l'offre aux clients d'une technologie de traitement parallèle et ils sont disponibles pour les applications sans qu'aucun changement ne soit requis pour l'application SQL.

IBM DB2 avec BLU Acceleration pour SAP Analytics

Comme nous l'avons précédemment vu, les charges de travail transactionnelles exigent un ensemble d'optimisations, centrées sur l'évolutivité et un débit maximal de transactions. Les charges de travail analytiques exigent un autre ensemble d'optimisations visant à offrir des temps de réponse très courts pour des requêtes vastes et complexes. Les technologies DB2 qui abordent les besoins des utilisateurs d'applications analytiques SAP, y compris SAP Business Warehouse, sont DB2 avec BLU Acceleration.

Accélération totale pour les analyses

IBM DB2 avec BLU Acceleration pour les analyses utilise une technique de compression allant au-delà d'une compression standard de style dictionnaire (à savoir, substituer un jeton pour chaque valeur et stocker chaque valeur précisément une fois sous forme compressée selon la gamme de valeurs et de types de valeurs qui existent dans cette colonne) en établissant un ordre d'évaluation pour les jetons afin que les valeurs réelles n'aient pas à être référencées et en dimensionnant les jetons selon la fréquence des valeurs qu'ils représentent, où les valeurs les plus fréquentes correspondent au jeton le plus court (1 bit). Cela permet au DBMS de détenir beaucoup plus de données en mémoire que cela est possible pour la plupart des DBMS tout en améliorant la performance des requêtes.

La performance est en outre améliorée en utilisant le traitement des vecteurs, comme nous l'avons vu précédemment. Tandis que le traitement des vecteurs n'est pas unique à DB2 avec BLU Acceleration, l'association de cette technique avec une approche de compression incluant une organisation en colonnes, le fait de sauter des données et l'optimisation du cache est impérative.

La vitesse en mémoire sans les limites de la mémoire

Certains DBMS optimisés en mémoire sont conçus pour exécuter des requêtes uniquement pour les données en mémoire. Comme une grande partie des données dans une base de données définie est rarement consultée, ceci est lourd à gérer. IBM DB2 avec BLU Acceleration est conçu pour optimiser l'utilisation de la mémoire et du disque, afin que toute la base de données n'ait pas à être en mémoire afin de réaliser tout le bénéfice de la vitesse en mémoire. Aussi, certains DBMS optimisés en mémoire ne peuvent fonctionner que sur une table si l'intégralité de cette dernière est en mémoire. IBM DB2 avec BLU Acceleration gère des données de table en s'assurant qu'une pleine vitesse optimisée en mémoire est atteinte sans que toutes les données demandées ne soient chargées en mémoire.

Optimisation pour IBM Power Systems

DB2 PureScale et DB2 avec BLU Acceleration sont conçus pour être utilisés sur x86 ou Power Systems. Cependant, pour un plus grand avantage, les concepteurs de DB2 ont collaboré avec les ingénieurs de matériels Power Systems pendant des années afin de garantir que Power Systems, notamment POWER7 et versions ultérieures, puisse pleinement profiter de DB2 et vice versa.

Les concepteurs ont optimisé avec précision l'environnement matériel/logiciel afin de s'assurer que le nombre immense de cœurs de Power Systems fonctionnent à leur capacité maximale pour répondre à la requête de BLU Acceleration à la vitesse exigée par votre activité. Conçu sur un héritage de forte résilience, disponibilité et sécurité, Power Systems, avec la toute nouvelle technologie sur processeur POWER8, offre des tailles de cache plus importantes, une bande passante mémoire supérieure et davantage de parallélisme (jusqu'à 12 cœurs par connecteur, avec 8 threads simultanés par coeur) afin de répondre à vos charges de travail analytiques et transactionnelles les plus exigeantes. La conception équilibrée de POWER8 optimise l'utilisation des cœurs et réduit ou élimine les temps d'attente des cœurs de processeur, entraînant des temps plus courts entre la requête et la solution.

Avantages de l'optimisation logicielle/matérielle

La mise au point de DB2 avec BLU Acceleration pour Power Systems signifie que les entreprises avec Power Systems peuvent exploiter leur infrastructure matérielle actuelle pour ajouter des analyses accélérées par BLU à leurs magasins de données SAP. Cela peut aider les sociétés à ajouter plus d'analyses sans étendre l'empreinte des infrastructures de serveurs.

Non disruptif pour un environnement IBM Power existant

Outre les avantages soulignés précédemment, il faut noter que, contrairement aux autres options de gestion des données d'application SAP, l'approche se basant sur DB2 s'adapte parfaitement à un environnement qui se base déjà sur la technologie IBM. Par exemple, si l'on prend l'exemple de la plate-forme IBM Power, ce choix est parfaitement cohérent, car il est possible de gérer globalement l'intégralité de l'environnement, plutôt que de traiter les ressources d'application SAP comme un cas spécial. BLU Acceleration n'est pas simplement un additif. Un support pour cette feature est intégré dans le noyau et les couches de gestion du stockage, ce qui

signifie que l'administration d'une base de données sur DB2 avec BLU Acceleration est simple pour les DB2 DBA. Aussi, DB2 avec BLU Acceleration offre une efficacité et une vitesse de base de données optimisée en mémoire sans que l'utilisateur n'ait à passer à une autre plate-forme, telle que Intel Xeon, comme pourrait l'exiger une autre approche.

Une alternative rentable et sans compromis à d'autres options de DBMS mixtes

Certaines approches DBMS qui supportent les applications SAP analytiques et transactionnelles impliquent une technologie émergente qui évolue rapidement et requiert une expertise spéciale pour la gestion. D'autres s'avèrent hors de prix par rapport au DB2 sur la plate-forme Power Systems et requièrent une mise au point importante. Non seulement le DB2 sur la plate-forme Power Systems permet une administration facile, de la vitesse et une évolutivité, mais il s'agit aussi d'une technologie éprouvée prise en charge par une vaste communauté de talents dans le monde.

Conforme à la direction de la future plate-forme de base de données d'IBM

Le fait d'utiliser DB2 sur Power Systems permet également de s'assurer que le centre de données est bien aligné avec la direction de la future plate-forme de base de données d'IBM. En supposant que d'autres applications et bases de données soient impliquées avec SAP, la capacité à coordonner les opérations des bases de données et à se concentrer sur les compétences de DB2 constitue un important avantage pour la future interopérabilité et la prise de décision.

Outre la capacité à lier ensemble différentes bases de données DB2, IBM offre un grand nombre d'options pour la coordination des bases de données utilisant des produits de la gamme InfoSphere.

PERSPECTIVES D'AVENIR

Il ne fait aucun doute que la technologie optimisée en mémoire représente le futur du DBMS. De nombreux fournisseurs possèdent aujourd'hui une certaine forme de cette technologie, et un plus grand nombre encore disposent de plans pour la faire connaître rapidement. Cependant, peu conviennent aux plate-formes de données pour les applications SAP. Certains d'entre eux peuvent actuellement fonctionner sur un matériel IBM mais peuvent ne pas représenter le meilleur choix pour un tel futur environnement. Il est certain qu'IBM DB2 sur Power Systems constituera toujours un choix optimal.

DIFFICULTÉS ET OPPORTUNITÉS

Bien que la technologie optimisée en mémoire d'IBM, y compris sa compression, son traitement des vecteurs et l'organisation en colonnes, fasse actuellement partie des technologies de pointe de l'industrie, de nombreux grands fournisseurs de DBMS continuent à développer leur technologie. IBM doit continuer à innover dans ces domaines afin de maintenir sa compétitivité.

CONCLUSION

Le monde de l'informatique subit des changements radicaux, motivés par des opportunités commerciales et des pressions concurrentielles, et rendus possibles par des changements en profondeur de la puissance et le coût réduit des processeurs et de la mémoire. Les tendances dans des domaines tels que les big data, l'informatique mobile, les médias sociaux et l'e-commerce à satisfaction instantanée créent un besoin supplémentaire en applications pour prendre en charge des analyses rapides pour de vastes recueils de données, pour le traitement intelligent de transactions extrêmes et des analyses en temps réel. SAP répond à ces pressions en faisant évoluer ses applications de façon appropriée. Le DBMS qui prend en charge ces applications doit agir de même.

La plate-forme DBMS qui parvient à faire cela doit satisfaire à ces principales exigences :

- Offrir des niveaux de performance pouvant être atteints uniquement grâce à une architecture de traitement parallèle optimale, augmentés par l'évolutivité flexible d'un groupe de stockage partagé.
- Prendre en charge des requêtes plus complexes avec moins de frais administratifs tout en entraînant une plus grande réactivité, ce qui est possible uniquement grâce à une organisation compressée en colonnes avec une exécution optimale par des techniques telles que le traitement de vecteurs.
- Permettre la coordination de données transactionnelles et analytiques en temps réel ou presque, de manière simple et facile à gérer.
- S'adapter parfaitement dans l'organisation stratégique globale du centre de données, sans introduire de systèmes physiques exotiques nécessitant une attention particulière, implémentant ainsi des systèmes pouvant être gérés dans le cadre d'une approche générale de gestion des systèmes et du stockage.

DB2, avec ses deux configurations (DB2 PureScale, pour le traitement des transactions évolutives, et DB2 avec BLU Acceleration, pour les analyses en temps réel), répond à toutes ces exigences. Lorsque vous ajoutez la dimension Power Systems, vous incluez une technologie conçue pour augmenter encore la performance de DB2.

Pour les utilisateurs SAP cherchant une plate-forme principale de persistance des données, IBM DB2 PureScale associé à IBM DB2 avec BLU Acceleration sur Power Systems répond à toutes ces exigences. Pour les utilisateurs SAP cherchant une capacité DBMS évolutive, robuste et éprouvée avec toute la vitesse du traitement en mémoire, IBM DB2 avec BLU Acceleration sur Power Systems constitue une bonne alternative à toute autre approche disponible.

À propos d'IDC

International Data Corporation (IDC) est le premier fournisseur international d'informations commerciales, de services de conseils et d'événements pour les marchés de la technologie de l'information, des télécommunications et de la technologie grand public. IDC aide les professionnels TI, les dirigeants et la communauté des investisseurs à prendre des décisions se basant sur des faits quant aux achats de technologie et à la stratégie commerciale. Plus de 1 100 analystes IDC proposent leur expertise locale, régionale et internationale en matière d'opportunités et de tendances de technologie et d'industrie dans plus de 110 pays dans le monde. Depuis 50 ans, IDC propose des points de vue stratégiques pour aider nos clients à atteindre leurs principaux objectifs commerciaux. IDC est une filiale d'IDG, plus grande société d'événements, de recherche et de moyens technologiques au monde.

Siège social international

5 Speen Street
Framingham, MA 01701
États-Unis
508.872.8200
Twitter : @IDC
idc-insights-community.com
www.idc.com

Avis de droits d'auteur

Publication externe des données et informations d'IDC : toute information d'IDC devant être utilisée dans des publicités, communiqués de presse ou matériels promotionnels requiert une autorisation écrite préalable de la part du Directeur national ou du Vice-président d'IDC concerné. Une ébauche du document proposé doit accompagner une telle demande. IDC se réserve le droit de refuser l'autorisation d'utilisation externe sans motif.

Copyright 2014 IDC. La reproduction sans autorisation écrite est totalement interdite.

Ce document a été développé grâce au financement d'IBM. Bien que le document puisse utiliser des documents publics issus de divers fournisseurs, dont IBM, il ne reflète pas nécessairement les positions de ces fournisseurs sur les problèmes abordés dans ce document.

