

ENTRETIEN AVEC CONOR O'MAHONEY

Eric Green

Bonjour. Je suis heureux de vous présenter une nouvelle série de baladodiffusions produites par le groupe Logiciels IBM qui étudie sous tous leurs angles les défis que les directions des TI et le personnel professionnel ont à relever aujourd'hui. Je m'appelle Eric Green et je m'entretiendrai avec des experts pour explorer de nouveaux exemples, perspectives et approches susceptibles de vous aider à relever ces défis et à vous présenter les possibilités des logiciels plus intelligents d'IBM. Allons-y.

Nous nous retrouvons cette fois-ci pour un nouvel épisode axé sur la gestion des données. La valeur et la disponibilité des systèmes, comme le besoin d'organiser les données, deviennent des éléments de plus en plus critiques au fur et à mesure que les organisations traitent de plus en plus d'informations. Pour parler de cet aspect des choses et de bien d'autres, nous nous entretenons aujourd'hui avec Conor O'Mahoney, Gestion d'informations, IBM, que nous remercions pour sa présence.

Conor O'Mahoney:

Merci. C'est avec plaisir que je participe à cet entretien.

Eric Green

J'ai pensé commencer en vous demandant comment la gestion des données a évolué au cours des dernières années.

Conor O'Mahoney:

Les deux ou trois dernières années ont été vraiment fascinantes en ce qui concerne la gestion des données et, parmi les points saillants, on trouve des systèmes matériels et logiciels intégrés. Des outils qui ont saisi une bonne partie des produits-phares d'IBM et d'Oracle, et qui sont à l'origine d'offres très intéressantes comme le système d'analytique plus intelligente d'IBM, le logiciel IBM PureScale Application System, Oracle ExoData et Oracle Exologic. Par la suite, d'autres fournisseurs ont suivi. Vous savez, un autre événement récent est l'émergence de ce que le groupe Forrester appelle la couche de compatibilité des bases de données. Il s'agit d'un ensemble de fonctions qui permettent aux utilisateurs de passer facilement d'un fournisseur de bases de données à un autre. Par exemple, en 2010, plus de 1 000 utilisateurs de bases de données Oracle sont passés à IBM DB2. Mais, ce qu'il y a de plus passionnant actuellement, ce sont les volumes massifs de données. Nous pouvons décrire cela comme un ensemble de défis de gestion des données trop volumineuses pour des méthodes plus traditionnelles comme les systèmes de gestion des bases de données relationnelles.

Eric Green

Tout cela est très intéressant et nous pourrions probablement avoir une conversation sur chacun de ces sujets, mais je me demandais si nous pouvions parler un peu plus des volumes massifs de données. Ce que je veux dire, c'est qu'il semble que les approches courantes des organisations sont limitées, et je me demande quel rôle ces nouvelles solutions joueront.

Conor O'Mahoney:

Eh bien! Il n'y a pas si longtemps, j'utilisais l'expression «trop gros». Ce qui, généralement fait référence au volume des données ou à leur débit. Par volume, j'entends qu'il y a simplement trop de données à gérer de façon fiable ou performante en utilisant des systèmes traditionnels. Vous savez, très bientôt les systèmes traitant des pétaoctets de données seront plus courants que rares. Certaines personnes peuvent avoir de la difficulté à imaginer que de tels systèmes pourraient se retrouver dans leur environnement. Mais elles ne devraient pas. Il s'agit simplement de revenir cinq ans en arrière, et de comparer les volumes de données de cette époque avec ceux d'aujourd'hui. En faisant cela, on réalise rapidement que les volumes de données croissent à un rythme incroyable, ce que cette cadence ne fera que s'accélérer au fur et à mesure que des points de données sont générés et entrent dans les systèmes. Quant à l'autre aspect de «trop gros», soit la vitesse, en fait, il s'agit probablement du plus important en regard des volumes massifs de données d'aujourd'hui. Par vitesse, j'entends que les données sont générées trop rapidement pour être traitées à temps en utilisant les systèmes conventionnels. L'expression clé ici, c'est «à temps». Il y a une certaine catégorie de données qui sont générées à une cadence tellement rapide qu'il est impossible de les traiter en temps réel ou quasi réel.

avec des systèmes traditionnels. Alors, à la multiplication des données générées par des machines, il faut ajouter la collecte à l'échelle du Web de données qui sont les premières responsables de ces déluges de données. Après la vitesse, on peut également utiliser un autre mot commençant par V pour qualifier ces données, et c'est variété. Les volumes de données massifs comprennent une variété de sources, dont les données structurées, documents non structurés, données verrouillées Web, relevés de capteurs en continu, données d'images et ainsi de suite. En résumé, les solutions pour volumes massifs de données prennent en charge les contraintes de volume, vitesse et de variété des systèmes traditionnels.

Eric Green

Pour développer cette idée, lorsqu'il s'agit d'une augmentation importante de données à traiter et ces entreprises voient vraiment croître leurs données, et bien d'où viennent ces données? Est-ce que ce phénomène va se poursuivre dans cette direction? Est-ce que l'on va voir une consolidation à un certain point, qu'est-ce qui favorise cette demande? Qu'est-ce qui explique une telle croissance de données?

Conor O'Mahoney:

Il y a quelques stimulants ici. Un de ceux-ci, c'est que de nombreuses organisations sont plus raffinées dans l'utilisation des données pour leurs prises de décision. Dans le passé, elles ont stocké beaucoup d'informations, mais ne les ont pas nécessairement exploitées en vue de prendre des décisions. Alors qu'aujourd'hui, ces organisations veulent tirer parti de toutes leurs informations. Ceci est donc un facteur, soit tout simplement faire intervenir plus de données dans la prise de décision. Un autre facteur, c'est qu'il y a beaucoup plus d'environnements dans lesquels des données sont générées par des machines. Il pourrait s'agir d'un environnement avec des compteurs intelligents. Vous savez, auparavant, votre compagnie d'énergie envoyait quelqu'un relever votre compteur une fois par mois. Ceci ne correspond pas à beaucoup de données, un relevé de compteur par mois. Mais avec les compteurs intelligents, ceux-ci communiquent avec une installation principale une fois par jour. Alors, déjà on voit que la quantité de données a été multipliée par 30. Et si ces compteurs communiquaient chaque 15 minutes avec l'installation principale? Tout se multiplie donc très vite, et les compteurs intelligents ne sont qu'un exemple. Il y a de nombreux autres exemples de compteurs ou capteurs ou systèmes de journalisation générant une quantité phénoménale de données.

Eric Green

Excellent. Merci beaucoup. Pendant que nous parlons d'exemples, pourquoi ne pas en profiter pour approfondir certains exemples que vous avez vus en temps réel chez IBM, comme les volumes massifs de données et les cas de réussite d'organisations qui ont vraiment dû déployer une bonne stratégie de gestion des données.

Conor O'Mahoney:

Certainement. OK. Voici un exemple d'un fournisseur de télécommunications. IBM collaborait avec lui pour offrir des promotions à ses abonnés. Ce fournisseur devait traiter plus de 100 000 dossiers à la seconde avec un délai de prise de décision de 10 millisecondes. Vous pouvez imaginer qu'avant les solutions dédiées aux volumes massifs de données, l'analyse de tels volumes à une telle vitesse n'était pas pratique. Un autre exemple, c'est celui d'un système intelligent de la circulation. Un tel système doit traiter plus de 250 000 sondes GPS et 630 000 segments à la seconde. Ici aussi, il y a un volume et une vitesse phénoménales de données générées par machine. Un autre exemple, c'est celui d'une agence qui doit traiter 600 000 dossiers à la seconde avec un délai de 1 à 2 millisecondes pour la prise de décision. C'est tout simplement ahurissant. Et enfin, le dernier exemple et non le moindre, et l'un de mes préférés – IBM Watson. Vous savez, ce superordinateur qui a été utilisé sur le jeu télévisé Jeopardy aux États-Unis. Ce système traite une variété, ce mot en V de notre définition originale, de données structurées et non structurées avec une latence de moins de 3 secondes pour la prise de décisions très complexes.

Eric Green

Pour poursuivre cette conversation, pourriez-vous nous parler un peu des technologies que les organisations peuvent mettre en pratique?

Conor O'Mahoney:

Absolument. Lorsque les gens parlent des technologies liées aux volumes massifs de données, ils peuvent faire référence à différentes technologies, dont les systèmes générés par cartes à base de grilles tels Apache Hadoop ou

IBM Big Insights. Ces systèmes peuvent traiter de grands volumes de données avec de faibles temps de latence. Les gens peuvent également faire référence aux systèmes qui analysent des données en continu, comme IBM Infosphere Streams. Ces systèmes traitent des volumes extrêmes de données avec des latences faibles. Ils peuvent également faire référence à des systèmes de traitement massivement parallèles comme IBM Netezza qui traitent de grands volumes de données avec des latences faibles. Une autre technologie courante dans ce domaine, ce sont les systèmes de base de données en mémoire comme IBM Solid DB qui traitent des données avec des latences vraiment faibles. Alors on peut voir qu'il y a différentes technologies comme les systèmes générés par cartes aux systèmes en continu, systèmes MPP, sans oublier les systèmes en mémoire et autres.

Eric Green:

Alors tout ce qui se dit sur les volumes massifs de données, c'est un peu comme si on signifiait la fin des systèmes de bases de données traditionnels tels que nous les connaissons. Est-ce la réalité qu'il faut comprendre?

Conor O'Mahoney:

Vous savez, il s'agit d'une excellente question. C'est une perception erronée de croire qu'il s'agit d'une situation binaire dans laquelle tout est blanc ou noir. Mais la réalité, et on peut la voir en travaillant avec ce type de solutions pour nos clients, c'est que de nombreuses situations réelles utilisent les deux approches ensemble. Ainsi, les technologies liées aux volumes massifs de données traitent la partie pour laquelle elles sont le mieux adaptées et les systèmes relationnels font de même. Il y a quelques instants, j'ai décrit des systèmes de volumes massifs de données dans le monde réel. Et bien, un des aspects vraiment intéressants de ces systèmes, c'est qu'ils ont tous une base de données relationnelle comme partie de la solution globale. Ainsi, la solution utilisée chez le fournisseur de télécommunications comprend une composante d'analyse en continu de concert avec un entrepôt de données relationnelles. Alors que le système intelligent de circulation et l'agence dont je parlais précédemment, utilisent tous les deux des données provenant de systèmes relationnels traditionnels dans la prise de décision. Et IBM Watson utilise une DBII pour travailler rapidement avec les réponses générées par la composante de volumes massifs de données. Alors, attendez-vous à ce que les systèmes relationnels traditionnels et les systèmes de volumes massifs de données travaillent ensemble dans vos centres informatiques. Et évidemment, un aspect à approfondir lorsque vous utilisez des solutions pour volumes massifs de données, c'est leur intégration avec des systèmes relationnels existants.

Eric Green

Très intéressant. Nous risquons de manquer de temps, mais il me semble que si vous recourez à une gestion de bases de données efficace, il y a des économies en argent à réaliser, de même que des économies en ressources, et l'expérience client est également touchée. Donc de nombreuses parties de l'entreprise sont affectées par ceci. En terminant, pourriez-vous nous donner vos deux, trois ou quatre meilleures suggestions sur ce que les organisations devraient prendre en compte lorsqu'elles gèrent leurs données.

Conor O'Mahoney:

Et bien, je crois que le point le plus critique pour tous aujourd'hui, c'est le contrôle des coûts. La gestion des données, en général, c'est de faire plus avec moins, et de maintenir les coûts. Il y a différentes stratégies pour ce faire. Une de celles-ci, c'est d'utiliser des systèmes de gestion de bases de données qui sont moins chères. Une autre, c'est de maximiser les technologies comme la compression des données; ceci peut vraiment permettre de faire d'importantes économies pour le stockage des données. Et une autre stratégie, c'est de bien connaître les coûts en personnel liés à certains produits. Vous savez, la dotation en personnel entraîne des coûts importants dans de nombreux environnements TI, et évaluer les ressources en personnel exigées pour certains produits est une mesure fort sage pour bien gérer le tout. Car en fait, non seulement vous devez composer avec des ressources en personnel, mais également avec la quiétude de ces ressources.

Eric Green

Excellent. Merci beaucoup pour ces informations. Et merci beaucoup Conor, d'avoir participé à notre entretien d'aujourd'hui.

Conor O'Mahoney:

Ce fut pour moi un grand plaisir, merci de m'avoir invité.

Eric Green

Merci d'avoir écouté. N'hésitez pas à visiter le site [IBM.com/software/ca/fr/](https://www.ibm.com/software/ca/fr/) pour entrer en contact avec des experts, poursuivre cet entretien ou obtenir des informations sur les logiciels plus intelligents proposés par IBM. Bâtissons une planète plus intelligente.