

Big data, BI : IBM dévoile DB2 10 et InfoSphere Warehouse 10

La base de données DB2 d'IBM n'avait pas connu de mise à jour majeure depuis 2008. Ce 3 avril, Big Blue a publié un communiqué et des informations dévoilant DB2 10 en même temps qu'InfoSphere Warehouse 10, lequel fonctionne sur DB2 et «s'intègre avec les systèmes dits 'Big data', comprime automatiquement les données pour réduire les flux à stocker et range les informations entre le passé et le futur, afin d'éliminer des coûts de code coûteux».

Les nouvelles versions sont le fruit de travaux des IBM Labs (IBM Research). Elles visent à «libérer du temps aux exploitants et utilisateurs de bases de données pour leur permettre de se consacrer à des tâches de plus forte valeur, comme le Big data ou le Business analytics [dont la BI, Business intelligence].»

Pour mener à bien le développement de nouvelles fonctionnalités, au cours des 4 années écoulées, IBM dit avoir collaboré avec plus de 100 clients, 200 partenaires IT et des centaines d'experts des IBM Research & Software Development Labs.

L'un des objectifs a été, par exemple, de pouvoir donner un accès à des nouveaux types de volumes de données colossaux, provenant des réseaux sociaux, des réseaux de capteurs (RFID), des équipements mobiles, tout autant que la prolifération des données métiers.

«Les nouveaux logiciels [version 10] aident les utilisateurs à accélérer par un facteur 10 la réponse à leurs requêtes et à libérer des capacités de stockage, jusqu'à un taux de 90% », explique IBM.

A priori, l'annonce de ce 3 avril concerne les versions Linux, Unix et Windows de DB2.

Dans sa communication, IBM cite le témoignage de Coca-Cola Bottling Company qui aurait économisé 1 million de dollars sur 4 ans -en licence, en maintenance et en coûts de stockage – grâce à sa migration vers DB2 9 et DB 10, à partir d'Oracle.

Les temps de requête auraient été considérablement diminués et les process de 'supply chain' très améliorés avec notamment des temps de réponse des traitements SAP accélérés de 30 à 60%.

Un responsable Datacenter de General Motors témoigne également dans ces termes:

«IBM et Intel ont collaboré sur toute la décennie écoulée pour optimiser DB2 avec la suite de développement Parallel Studio 2011 d'Intel sur des processeurs Xeon. Nous sommes très enthousiastes à l'idée de vérifier des améliorations dans un facteur 10 dans les requêtes, sur nos systèmes x3850 sur Xeon E7.»

Les principales nouveautés



. **Adaptive Compression** : cette fonction comprime instantanément les données;

. **Multi-Temperature Data Management** permet de notifier avec quelle fréquence les données sont

utilisées et nécessaires, ce qui permet de les ranger par catégories d'unités de stockage « chaudes » ou « froides », coûteuses ou moins coûteuses;

. **Time Travel Query** : le statut 'bi-temporel' entre passé et futur permet d'améliorer les prises de décisions et de détecter des incohérences de calendrier (dans des réservations de transports et d'hôtels). Cette fonction, qui travaille aussi bien en mode transactionnel qu'en temps réel ou 'valid time', permet également de créer très rapidement des rapports 'historiques'

. **compatibilité** améliorée avec **Oracle** : DB2 10 et InfoSphere Warehouse 10 apportent un support natif des données de syntaxes SQL, PL/SQL («*compatibilité avec le code supérieure à 98%*») ou des 'scripts' ou des packages intégrés aussi bien que des modèles concurrents des bases Oracles -«*ce qui facilite le passage de bases Oracle à DB2. Les utilisateurs d'Oracle se retrouveront ainsi plus facilement dans DB2 et InfoSphere Warehouse*».

. le support de l'environnement PureScale dans DB2 est élargi, amélioré et facilité;

. un utilitaire « **parallel continuous ingest** » est ajouté, qui permet l'injection de données en continu, «*une fonction importante pour le data warehousing ou la BI en temps réel*».

. les fonctionnalités NoSQL Graph Store (permettant de travailler des 'graph triples') sont ajoutées ainsi que le langage de requête SPARQL.

D'autres améliorations portent sur la sécurité, le load-balancing, la gestion des index, le monitoring, la haute disponibilité et la résilience des données et des bases.