

Evaluation des solutions d'intégration de données en temps réel

Sommaire

- 2 Introduction**
- 2 L'exigence d'une intégration en temps réel**
- 4 Pourquoi les outils ETL traditionnels ne sont pas toujours la solution**
- 5 La technologie Change Data Capture et l'intégration de données en temps réel**
- 7 A propos d'IBM InfoSphere Change Data Capture**
- 8 Les Bénéfices d'IBM InfoSphere Change Data Capture**
- 9 Qui a besoin d'IBM InfoSphere Change Data Capture?**
- 10 Une solution temps réel pour des problèmes temps réel**
- 10 Reporting et analyse**

Introduction

L'accès à l'information immédiate et à jour est un facteur clé de succès. Les sociétés ont besoin d'accéder à l'information précise pour prendre de meilleures décisions, mieux satisfaire les clients et réagir aux changements se produisant sur leurs marchés.

Cependant, les sociétés se heurtent de plus en plus à des problèmes dans leur recherche d'une information précise et à jour. La plupart des grandes PME et des grands comptes disposent aujourd'hui d'un datawarehouse capable d'analyser l'information depuis des systèmes de production sur une base de données quotidienne ou hebdomadaire, cependant la plupart jugent le processus d'extraction de l'information trop long et complexe, vu qu'il se fait sur des bases de données critiques sur employées. En même temps, le besoin d'accéder à l'information à jour a augmenté au point que les mises à jour quotidiennes ou hebdomadaires ne suffisent plus.

Ce document examinera les challenges commerciaux conduisant les entreprises à considérer l'intégration de données temps réel ainsi que les forces et faiblesses des différentes approches répondant à ces problématiques. Il fournira aussi un aperçu des solutions d'intégration de données à faible impact et temps réel.

L'exigence d'une intégration en temps réel

Qu'est ce qui maintient les directeurs des systèmes d'information éveillés ? Probablement plein de choses, y compris les projets d'intégrer toutes les données. Selon une étude Forrester réalisée auprès de 600 décideurs de solutions technologiques travaillant pour des entreprises basées en Amérique du Nord, améliorer l'intégration entre applications est une des dix priorités pour les entreprises en 2008.

Pourquoi ? Regardons quelques - uns des défis commerciaux actuels :

- Une exigence d'information temps réel accrue pour le reporting et l'analyse. En général, le reporting s'effectue depuis les datawarehouses qui sont mis à jour sur une base quotidienne ou hebdomadaire. Pour la plupart des rapports, ces données sont suffisamment à jour. Pour d'autres, cependant, seule une information à la minute conviendra, c'est le cas des données de stocks où le produit en stock peut être en grosse ou faible quantité, ou lorsqu'il existe des besoins de facturation à la minute ou à la demi-journée.
- Les grands volumes d'informations sont difficiles à gérer dans une fenêtre batch. Plus l'information collectée est importante – comme les données relatives aux transactions en ligne, les données de stocks, l'information client, plus l'effort pour déplacer cette information vers le datawarehouse augmente. De nombreuses sociétés considèrent qu'une fenêtre batch de huit heures n'est plus suffisante pour permettre aux outils ETL traditionnels (extraction, transformation, chargement) d'intégrer toutes les données nécessaires.
- Nécessité de conduire des opérations 24 heures sur 24, 7 jours sur 7, pour réduire les fenêtres batch. Comme les transactions se font dans les pays du monde entier et sur le web, de nombreuses sociétés rencontrent un problème de lenteur lié aux fenêtres batch, rendant les outils ETL traditionnels incapables d'extraire les données en un temps limité.
- Un besoin croissant de détecter et de réagir aux événements commerciaux dès qu'ils se produisent. De nombreuses sociétés recherchent des moyens pour détecter les événements commerciaux en provenance des systèmes de production et obtenir ces événements en déclenchant une réponse dans un autre système. Par exemple, une société vendant des téléphones portables souhaite envoyer un message texte à un client n'ayant presque plus de forfait pour lui proposer d'acheter de nouvelles minutes/heures.
- Le besoin de traquer tous les changements pour des besoins d'audit. Les sociétés ont besoin de se mettre en conformité avec les réglementations qui exigent des données avec les informations sur toutes leurs modifications et pas seulement le résultat de ces changements.

- Un besoin croissant d'obtenir les données synchronisées dans toute l'entreprise. Les clients veulent un accès à la minute aux commandes, paiements et données relatives aux stocks pour pouvoir acheter des produits, payer les factures et vérifier l'état de la livraison online. Les employés ont besoin d'à peu près les mêmes informations pour offrir un meilleur service clients et prendre des décisions commerciales avisées. Pour parvenir à ce résultat, les données d'e-commerce ont besoin d'être synchronisées avec les applications métier et les données utiles pour être transférées en temps réel dans toute l'entreprise.
- Les sociétés veulent déployer de nouvelles applications sur des applications existantes pour augmenter le débit de chargement. Souvent, les systèmes existants sont saturés et augmenter les capacités coûte très cher. Les sociétés veulent de nouvelles applications sur Unix ou Windows pour éviter les coûts mais intégrer les données depuis des systèmes existants sans augmenter le temps de chargement est un vrai challenge.

Pourquoi les outils ETL traditionnels ne sont pas toujours la solution

La majorité des entreprises utilisent des outils ETL pour extraire les volumes de données depuis leurs systèmes de production et les télécharger dans d'autres systèmes y compris les datawarehouses. La force principale des outils d'ETL est leur capacité à extraire les données depuis différentes applications, à exécuter des transformations complexes sur les données, plus leur chargement de gros volumes de données dans les datawarehouses.

Mais lorsque se présentent des besoins d'extraction de données depuis des systèmes de production, il y a plusieurs restrictions à l'utilisation d'un outil d'ETL. Ces limitations sont les suivantes :

- L'ETL est un processus batch exigeant une fenêtre batch – Les outils ETL peuvent déplacer des données batch depuis des systèmes de production vers des datawarehouses. Pour de nombreuses entreprises, exécuter ce processus batch lorsque les systèmes ne sont pas trop utilisés suffit. Cependant, avec la globalisation constante des marchés et l'utilisation croissante d'Internet, trouver une période où les systèmes de production peuvent être arrêtés pour exécuter une extraction des données en mode batch devient de plus en plus compliqué.

- Les fenêtres ETL batch augmentent avec les volumes de données extraits. Même lorsque les entreprises peuvent tolérer une fenêtre batch durant la nuit, la plupart considèrent qu'avec des volumes de données à extraire de plus en plus importants, la fenêtre doit faire l'extraction plus de huit à douze heures par jour (c'est-à-dire au-delà de la période durant laquelle les systèmes de production sont off line).
- Les outils ETL exigent que les changements soient effectués au niveau des systèmes de production depuis lesquels ils extraient les données. Les outils ETL exigent souvent de procéder à des changements sur les systèmes de production, par exemple, en ajoutant la date ou l'heure pour des mises à jour incrémentales. Pour de nombreuses sociétés, le risque additionnel que cela engendrerait sur leurs systèmes critiques n'est pas tolérable.
- Les outils ETL ne traquent pas tous les changements qui ont été faits dans le système ; ils traquent plutôt le résultat de ces changements. Par exemple, si le niveau de stocks d'un certain produit venait à diminuer suite à des achats mais ensuite à augmenter suite à un stockage, ou si le statut de la commande devait changer, les outils ETL seraient incapables de traquer ces changements.

La technologie Change Data Capture et l'intégration de données en temps réel

Change Data Capture (CDC) fournit une manière de capturer les changements sur les systèmes de production afin qu'ils puissent être appliqués ailleurs sans attaquer directement la base de données.

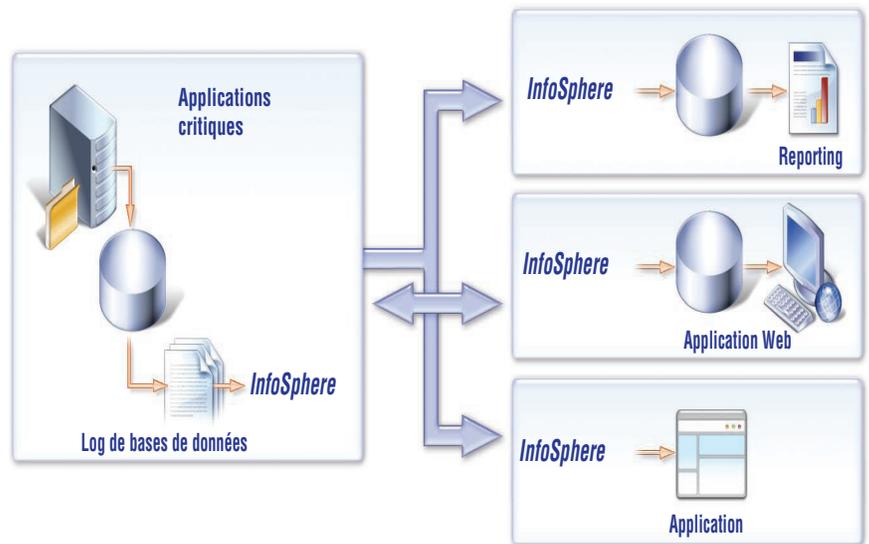
Les solutions utilisant la technologie CDC ont de nombreux avantages par rapport aux outils ETL traditionnels et les complètent dans de nombreux cas. Les avantages par rapport aux ETL sont les suivants :

- Les changements sont capturés en temps réel afin que l'information soit toujours à jour. La technologie CDC permet de capturer les changements dès qu'ils se produisent. Il en résulte une information toujours à jour plutôt qu'une information datant de la dernière exécution de la fenêtre batch.
- Pas d'impact sur la performance des systèmes de production. CDC lit les logs de bases de données et n'exécute pas des requêtes sur les bases de données directement. Il n'y a pas d'impact sur les systèmes de production.

- Pas d'exigence de fenêtre batch. Avec les changements capturés, transformés et appliqués en continu, il n'y a pas besoin d'arrêter les systèmes pour extraire des données.
- S'adapte facilement à de très grandes bases de données et à un grand nombre de transactions. Plutôt que de répliquer toutes les données, seuls les nouvelles données sont répliquées dans les tables. Il en résulte une plus grande souplesse avec des données déplacées moindres.
- Aucune modification n'est à apporter au système source. La technologie CDC ne fait que lire les logs des bases de données et n'impose aucune modification de la base de donnée source. Grâce à cette technologie, la totalité des transactions peut être détectée et on peut même obtenir l'historique des changements (utilisateur, application, heure, etc.).
- Capture tous les changements des systèmes et pas uniquement les résultats des changements. Pour l'audit et la mise en conformité, toutes les actions d'insertion, de mise à jour et de suppression sont enregistrés et pas uniquement le résultat de ces actions.
- Peut compléter les outils ETL traditionnels. De nombreuses entreprises combinent les forces du CDC aux outils ETL traditionnels. Par exemple, la technologie CDC permet d'accéder aux données directement depuis les bases de données et de répliquer les données depuis des systèmes opérationnels sans nécessiter de fenêtre batch ou d'interruption de systèmes critiques. Les données sont ensuite chargées dans des operational data stores (ODS), avec les outils ETL alimentant les datawarehouses ou les autres applications. Les systèmes de production peuvent continuer à fonctionner 24 heures sur 24 et le reporting temps réel peut être supporté par l'ODS.

A propos d'IBM InfoSphere Change Data Capture

IBM InfoSphere Change Data Capture fournit l'intégration bidirectionnelle et temps réel et les transformations entre diverses bases de données relationnelles et d'autres données stockées sur différentes plates-formes. Contrairement à des outils traditionnels ETL, des chargements batch et des requêtes et systèmes basés sur les messages, IBM InfoSphere Change Data Capture a été conçu pour fournir une intégration de données souple et hautement performante en temps réel avec une latence minimum. En effet, la technologie unique Change Data Capture d'IBM a un impact minimum sur la performance des applications opérationnelles.



REPORTING
Business Objects
Cognos
Panorama

E-BUSINESS
Microsoft .NET
IBM WebSphere
BEA WebLogic

DISTRIBUTION DE DONNEES ET CONSOLIDATION
DB2
Sybase
Message Queue
UDB
Oracle
Teradata
SQL Server
PointBase

Les bénéfices d'IBM InfoSphere Change Data Capture

Faible impact – Plutôt que d'utiliser des triggers ou d'exécuter des requêtes sur la base de données, IBM InfoSphere Change Data Capture utilise le meilleur de la technologie Change Data Capture (CDC) basée sur les logs, pour capturer les changements de bases de données depuis des logs de bases de données. Cela garantit que même les applications les plus critiques basées sur le système source ne sont pas malencontreusement impactées.

Facile à déployer – La facilité d'utilisation de l'interface GUI d'IBM InfoSphere Change Data Capture facilite la sélection de la source et des bases de données cible et configure ensuite les transformations. Le résultat est le suivant : implémentation plus rapide, coût moindre et retour sur investissement plus important.

Temps réel – Les changements de données sont capturés dans le système source au fur et à mesure que ceux-ci se produisent et les changements sont immédiatement distribués aux systèmes cible. Toute l'information est mise à jour et synchronisée.

Flexible – Les clients utilisent IBM InfoSphere Change Data Capture pour résoudre de nombreux problèmes métier sur toute une gamme de plateformes, depuis les mainframes jusqu'aux solutions pour mobiles. Utilisé pour charger une base de données, distribuer et synchroniser les données entre différentes applications, ou gérer d'autres besoins basés sur la réplication. Il peut être déployé sur de nombreuses architectures telles que DB2, UDB, Microsoft SQL Server, Oracle, Sybase, Teradata, et PointBase.

Adaptable – InfoSphere Change Data Capture est une solution hautement performante facilement adaptée à de gros volumes de données sans impacter la performance des applications critiques sur le système source.

Qui a besoin d'IBM InfoSphere Change Data Capture ?

- Les sociétés ayant besoin de charger leurs datawarehouses en temps réel afin de pouvoir prendre des décisions opérationnelles et tactiques basées sur la toute dernière information.
- Les sociétés ayant besoin de fournir à leurs clients un accès temps réel depuis leur site web tout en minimisant l'impact des systèmes gérant les applications logistiques qu'ils utilisent pour améliorer leurs opérations en ligne.
- Les sociétés ayant besoin d'offrir à leurs clients un accès temps réel avec l'information relative aux stocks depuis leur site web tout en minimisant l'impact des applications critiques qu'ils utilisent pour garantir des opérations efficaces en continu.
- Les sociétés ayant de nombreuses applications de production ou qui ont subi une fusion ou acquisition et ont besoin de consolider et de distribuer des données entre systèmes hétérogènes.
- Les sociétés cotées en bourse ayant besoin de mener des opérations de reporting financier depuis toute une gamme de systèmes transactionnels dans différentes régions, départements et unités commerciales.
- Les sociétés ayant besoin d'améliorer la performance opérationnelle de leurs systèmes, qui seraient autrement affectés par un ralentissement des process batch de nuit et de coûteuses opérations de requête et de reporting.

« La gestion de parcs est si complexe qu'en obtenant des informations synthétiques, toute une série de décisions clés en découlent dans les domaines de la maintenance, du help desk et du financement » affirme Mr Gilles De Lignac, Directeur informatique chez SCC.

Une solution temps réel pour des problèmes temps réel

Faciliter les projets d'e-Business

SCC, premier groupe informatique privé et acteur majeur européen des services informatiques d'infrastructure utilise IBM InfoSphere Change Data Capture pour répliquer et intégrer bi-directionnellement ses données hétérogènes (DB2, SQL...) afin, d'une part de centraliser et de consolider les données hétérogènes émanant de différents systèmes d'information de différents pays, d'autre part d'offrir aux clients un extranet temps réel (avec modification de leurs données) avec une visibilité en temps réel sur leurs informations stratégiques : commandes, livraisons, stocks, parc informatique...

Ainsi IBM a permis à SCC et à ses clients d'obtenir un réel avantage concurrentiel.

En obtenant par exemple les statistiques sur son parc, le client peut réaliser au plus vite les actions appropriées pour sa mise à niveau. C'est ce projet temps réel « Life Cycle » qui permet à SCC de se différencier fortement de la concurrence et d'offrir à ses clients un réel avantage concurrentiel – la société SCC signant d'ailleurs ainsi de nouveaux clients grâce à cet avantage temps réel clairement perceptible. Etre capable d'être réactif et de servir ses clients au plus vite, c'est un facteur clé de succès pour une société ayant une activité service importante.

Reporting et analyse

Depuis sa création en 1936, Ricoh développe des appareils de traitement de l'image et autres produits, services et solutions qui mettent l'accent sur la création de valeur afin de répondre en permanence aux besoins de ses clients. Son siège pour la région Europe, Moyen Orient et Afrique, est situé à Amstelveen aux Pays-Bas. Ses usines sont basées, l'une en France, l'autre en Angleterre.

Ricoh devait relever les défis suivants :

- Un nombre d'applications en augmentation
- Des volumes de données en augmentation
- Une capacité de traitement limitée
- Une fenêtre batch trop « serrée »
- Un problème de disponibilité du datawarehouse
- Des utilisateurs disposant d'une information « trop pauvre »
- Des besoins d'entreprise croissants pour un reporting temps réel

Grâce à IBM InfoSphere Change Data Capture, Ricoh Europe a pu :

- Sur le plan informatique : augmenter le nombre de ses applications, répondre aux problèmes des travaux batch non terminés à temps, diminuer les temps de chargement.
- Sur le plan commercial : être plus réactif et dynamique en obtenant de l'information de dernière minute.

La société utilise donc IBM InfoSphere Change Data Capture pour faire du reporting en temps réel et réduire les volumes de données et le temps d'exécution total.

Elle a obtenu les bénéfices suivants par rapport au datawarehouse :

- Réduction très importante du volume de données extrait
- Réduction très importante du temps pour extraire les données
- Données disponibles en temps réel dans l'ODS / Datawarehouse (IBM DB2 8 on Windows 3)
- Diminution de l'impact sur les systèmes sources

IBM InfoSphere Change Data Capture a été installé chez Ricoh Portugal et sera bientôt implémenté en Espagne, Pologne, Finlande et Hongrie.



Pour en savoir plus

Pour en savoir plus sur IBM InfoSphere Change Data Capture, contactez votre représentant commercial IBM ou votre partenaire commercial IBM. Consultez notre site Web à l'adresse suivante :

ibm.com/services/fr

© Copyright IBM Corporation 2008

Compagnie IBM France
Tour Descartes La Défense 5
2, avenue Gambetta
92066 Paris La Défense Cedex

Imprimé en France
04-08
Tous droits réservés

IBM et le logo IBM, IBM InfoSphere Change Data Capture, DB2, Websphere sont des marques d'International Business Machines Corporation aux États-unis et/ou dans certains autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques de Microsoft Corporation aux États-Unis et/ou dans certains autres pays.

Java et toutes les marques commerciales basées sur Java sont des marques commerciales de Sun Microsystems, Inc. aux États-Unis et/ou dans certains autres pays.

Les autres noms de société, de produit ou de service peuvent appartenir à des tiers.

Le fait que des produits ou des services IBM soient mentionnés dans le présent document ne signifie pas qu'IBM ait l'intention de les commercialiser dans tous les pays où elle exerce une activité.

Toutes les déclarations relatives à l'orientation future et aux intentions d'IBM sont sujettes à modification ou à retrait sans préavis et représentent uniquement des objectifs et des buts.