

Sohaib Abbasi, Informatica : « Une grande confusion règne sur le Big Data »

Silicon.fr : En quoi consiste la virtualisation de données mise en avant par Informatica ?

Sohaib Abbasi : Pour concevoir ses applications, le développeur sélectionne les bases de données directement dans son SGDB (Oracle, SQL Server, IMS, Hana...) et manipule les tables, les champs, etc. Il s'agit d'une approche traditionnelle d'applications transactionnelles sur site.

Aujourd'hui, les applications utilisent aussi les informations des médias sociaux, de l'Internet des objets et de diverses autres sources. De plus, elles sont déployées sur site et/ou via le cloud.

Avec la technologie de machine virtuelle de données Informatica Vibe, le mapping de données unique est généralisé et virtuellement déployé partout. D'où notre slogan : « Map One Deploy Everywhere ».

Pour comprendre ce concept, on peut comparer le mécanisme à Java. Il suffit qu'une application, un service ou un équipement intègre la machine virtuelle Java pour pouvoir exécuter du code Java. De même, en intégrant notre technologie Vibe, tout service ou équipement pourra interagir avec le mapping de données.

A quels services et produits existants peut s'appliquer cette solution ?

Vibe permet d'étendre le mapping aux systèmes transactionnels (Oracle, SAP, etc.), aux solutions SaaS (Salesforce, Workday...), au media sociaux et à l'Internet des objets. La connectivité est alors activée vers et depuis toutes ces informations. Vibe intervient comme l'esperanto des données, et le développeur n'a plus à se préoccuper ni de la localisation, ni de la spécificité technique de toutes ces données.

Dès à présent, cette technologie est intégrée à toutes les solutions Informatica, y compris les plus anciennes. Il suffit au client de télécharger la mise à jour pour disposer de ces technologies, dont une connectivité Hadoop entre autres.

Comment prévoyez-vous d'intégrer les autres acteurs du marché pour diffuser votre machine virtuelle de données ?

D'ici à quelques mois, Informatica proposera Vibe en tant que Software Development Kit (SDK), c'est-à-dire des APIs et extensions qui permettent à un éditeur ou une entreprise d'ajouter de nouveaux connecteurs et des fonctions de transformation. Ainsi tous les logiciels ou produits "Vibe Inside" (embarquant la machine virtuelle de données) pourront interagir et comprendre les mappings Vibe. Vibe devient le serveur intégré de qualité et d'intégration des données (accès, agrégation, gestion...), au même titre qu'un serveur applicatif exécute les applications.

Tout l'écosystème d'Informatica, et au-delà, pourra intégrer Vibe à ses produits ou concevoir des fonctions d'extension pour notre technologie.

Certaines entreprises utilisent déjà avec succès Vibe Inside, comme Dun & BradStreet.

A l'avenir, Vibe for Internet of Things intégrera aussi l'Internet des objets : capteurs et senseurs envoyant ou recevant des données dans le domaine du machine-to-machine (M2M).

Une nouvelle solution signifie-t-elle aussi de nouvelles compétences à acquérir ?

Jusqu'à présent, le développeur devait maîtriser des langages comme Cobol, Visual Basic, C++, Java, Ruby OnRails, Javascript... Mais aussi la manipulation des données, la virtualisation, le multitenant, etc.

Avec Vibe, le programmeur n'a plus besoin de savoir tout cela. Il a très peu de connaissances à acquérir et il devient jusqu'à dix fois plus productif. Et les utilisateurs d'Informatica sont déjà opérationnels.

Ainsi, une banque utilisant les services très chers de programmeurs MapReduce pour ses projets Hadoop a décidé de passer sous Vibe. Résultat : la solution a gagné 30 % de performances et ce, sans que les développeurs n'aient eu à connaître Hadoop.

Enfin, la gestion des multiples bases de données (IMS, Oracle, Hadoop/HDFS, Hana...) est remplacée par une administration globale via le mapping unique sous Vibe.

Que pensez-vous de l'engouement pour le Big Data ?

Une grande confusion règne sur le Big Data et l'informatique analytique, qui ne sauraient se résumer à Hadoop. Selon moi, on peut considérer l'analytique en six familles technologiques :

- Le datawarehouse (avec des sociétés comme Teradata) ;
- Les appliances (EMC/Greenplum, IBM Netezza...);
- Les technologies In-Memory comme SAP Hana ou Oracle In-Memory ;
- Les services cloud comme on en trouve sur Amazon ou Google ;
- Les solutions Hadoop comme Cloudera, MapR ou Hortonworks

De nombreux éditeurs affirment pouvoir couvrir tous ces aspects. Ce qui est rarement vrai, et surtout pas avec le même outil. Avec Vibe, il est possible d'intégrer et de gérer la qualité de tous ces types d'information.

Quelle valeur ajoutée métier apportent réellement le Big Data et Hadoop ?

Selon moi, le Big Data est à la confluence de trois disciplines : Big Transactional Data (les SGBD traditionnels), Big Interaction Data (interactions des médias sociaux et du M2M ou Internet des objets) et Big Data Processing (Hadoop).

Côté valeur ajoutée, les applications analytiques ont déjà fait leurs preuves.

Les interactions sur les médias sociaux combinées aux données relationnelles permettent de mettre en relief des tendances et observations souvent impossibles à déceler autrement (influence des réseaux sociaux sur les ventes, etc.). Les corrélations des informations du M2M permettent de concevoir de nouveaux services comme on le constate avec les opérateurs télécoms ou les services

de géolocalisation, par exemple.

Enfin, Hadoop a été créé par des spécialistes du Web pour accélérer les traitements et, surtout, réduire les coûts des serveurs nécessaires à ces traitements très lourds. C'est pourquoi Hadoop peut fonctionner sur des serveurs x86 standard. Cependant, il faut relativiser tout cela. En effet, les Yahoo, Facebook ou Google utilisent des versions souvent anciennes des briques Hadoop, tandis que Cloudera, MapR et Hortonworks proposent les dernières évolutions.

Voir aussi

[Silicon.fr étend son site dédié à l'emploi IT](#)

[Silicon.fr en direct sur les smartphones et tablettes](#)