

Vos données de référence se présentent sous la forme d'un graphe : Êtes-vous prêt ?

Karen Lopez, InfoAdvisors
Mars 2015



InfoAdvisors

www.datamodel.com
karenlopez@Infoadvisors.com
[@datachick](#)

*Ce livre blanc est sponsorisé par Neo Technology Inc.
Neo4j est une marque déposée de Neo Technology Inc*

HISTOIRE DE DONNÉES

Vos données souhaitent raconter une histoire. Ce n'est pas un *problème de données* ; c'est une *histoire de données*.

Vos données de référence se présentent sous la forme d'un graphe - Êtes-vous prêt ?

Gestion des données de référence aujourd'hui

Il est probable que votre programme de gestion des données de référence utilise la même technologie de base de données que votre application de transaction, à savoir une base de données relationnelle et mature dont les réglages sont optimums (SGBDR). Vous gérez parfaitement les bases de données relationnelles car cela fait des années que vous les utilisez et que la plupart de vos données sont gérées de cette façon. Il est donc compréhensible que vous conserviez ce mode de gestion pour vos données de référence. **Et si les relations de données que vous connaissez et qu'il vous reste encore à découvrir vous incitaient à les gérer exactement de la même façon que les données elles-mêmes ?** Ces histoires de données deviennent un avantage concurrentiel de plus en plus important alors que nous entrons dans une ère nouvelle d'analyse des données et des activités économiques.

En règle générale, la gestion des données de référence comprend les données relatives aux clients, aux produits, aux comptes, aux ressources, aux partenaires ainsi que tout autre type de données que l'on transmet facilement au sein d'une entreprise.

Compte tenu du nombre de trouvailles récentes concernant les approches de gestion des données, nous disposons d'une large variété de définitions de la gestion des données de référence. Ces définitions vont d'une définition très physique de la gestion de données dans un seul fichier ou une seule source de données, à une définition plus conceptuelle de la gestion des ressources de données, accessibles à partir d'un seul service ou d'une seule application. Cependant, dans les deux cas, nous mettons l'accent sur les données en tant que concept unique d'un seul type.

En tant que professionnels de la gestion des données, nous avons dû lutter pour que nos mises en œuvre traditionnelles des données de référence puissent faire face aux évolutions des activités et de l'environnement économiques. Comme les données de référence, par définition, sont communément diffusées, cette lutte se fait au détriment de la souplesse commerciale et affecte l'ensemble de la société. Nos architectures visaient à faire correspondre les données à une seule définition, ce qui constitue aujourd'hui une solution à court terme pour la plupart d'entre nous.

L'avenir de la gestion de données de référence

Les programmes de gestion des données de référence qui ont essayé de continuer à sauvegarder physiquement les données à un seul emplacement doivent encore faire face aux réalités de la technologie de l'information moderne. La plupart des entreprises utilisent des applications conçues par

*Vos données de référence se présentent sous la forme d'un graphe :
Êtes-vous prêt ?*

des fournisseurs : systèmes de gestion de la relation client (GRC), systèmes de gestion du travail, de la comptabilité fournisseurs, de la comptabilité clients et des points de vente, etc. Avec cette approche, il n'est pas toujours possible de transférer toutes les données de référence vers un seul emplacement. Même avec un système opérationnel de gestion de la relation client, les informations client sont souvent stockées dans plusieurs systèmes. Il en va de même pour les données relatives aux produits et à la comptabilité.

Les programmes les plus puissants n'auront aucun mal à trouver un emplacement physique unique pour toutes les données. Ils fourniront également les normes, outils et services nécessaires pour fournir une vision cohérente des données d'entreprise. Certaines données pourront être stockées à un même emplacement, à l'aide des technologies correspondant le mieux à leur histoire. On pourra également trouver des données dans plusieurs systèmes physiques grâce à l'utilisation croissante d'applications intégrées et aux besoins en termes de performances et de traitement géographiquement distribué. Une fois que nous comprenons notre environnement, nous pouvons concevoir des solutions en fonction des besoins associés.

Les programmes les plus puissants n'auront aucun mal à trouver un emplacement physique unique pour toutes les données relatives à un concept. Ils fourniront également les normes, outils et services nécessaires pour fournir une vision cohérente des données d'entreprise.

L'avenir de la gestion de données de référence réside dans l'exploitation des données et de leurs relations entre elles. La gestion des données de référence consistera à en fournir des analyses cohérentes et significatives. Dans de nombreux cas, nous serons en mesure de rassembler les données en un emplacement unique, notamment pour optimiser les performances des requêtes et l'ajustement des données. Les bases de données de graphes offrent ce type d'ajustement des données et des performances, comme décrit ci-dessous. Dans ce livre blanc, nous expliquons en quoi vos données de référence sont un graphe et comment les bases de données de graphes comme Neo4j constituent la meilleure technologie pour les données de référence.

Comment les relations de données offrent-elles de meilleures perspectives ?

Lorsque nous pensons aux relations de données, ce sont souvent des clés étrangères qui nous viennent à l'esprit. Leurs fonctions de base de données relationnelles nous permettent de relier des lignes de données entre elles. Et il ne s'agit là que d'un type de relation. **Du point de vue technique, ces clés étrangères constituent de simples contraintes pour vos données.** La relation de clé étrangère entre OPÉRATION et ÉLÉMENTS DE LIGNE D'OPÉRATION permet de s'assurer (au moyen d'une contrainte) que tout ÉLÉMENT DE LIGNE D'OPÉRATION dispose bien d'une OPÉRATION correspondante. Les meilleures histoires de données mettent l'accent sur des questions beaucoup plus complexes et polyvalentes :

- Quels produits sont généralement achetés ensemble ?
- Quel client et quels membres de son foyer ont davantage tendance à acheter tel type de produit régulièrement ? Qu'en est-il des groupes de clients ?
- Quels produits et services doit-on recommander aux clients ? Doit-on les proposer séparément ou dans le cadre d'une offre groupée ?
- Quels sont les autres facteurs qui influent sur les décisions prises par les clients au moment de l'achat ?

Nous évoquerons davantage de questions de ce type ultérieurement.

Graphes : Théorie et pratique

Les graphes sont des structures de données qui décrivent à la fois les données et leurs relations. Les formes de graphe les plus souvent reconnues sont les réseaux et les hiérarchies. Les graphes comportent des nœuds entre lesquels existent des relations. Il est possible de définir des propriétés pour les nœuds et les relations. Des étiquettes peuvent également être appliquées aux nœuds.

Vos données de référence se présentent sous la forme d'un graphe :
Êtes-vous prêt ?

Réseaux

Un réseau comprend un ensemble de nœuds ainsi que les relations entre eux.

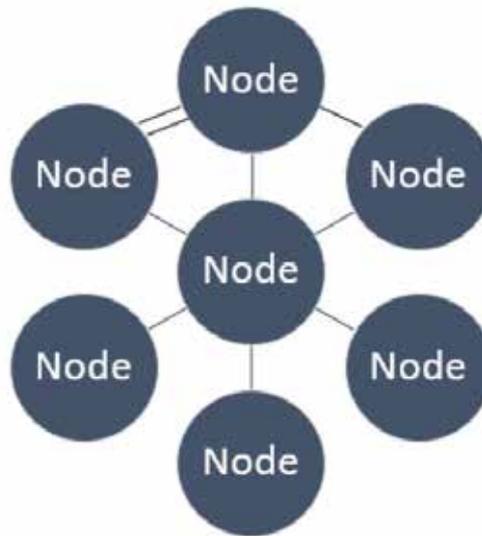


Figure 1 – Réseau

Vous pouvez penser à un réseau d'amis, de serveurs, de bases de données ou de clients. Dans une bases de données de graphes, il n'est pas nécessaire que les nœuds soient tous du même type. La même règle s'applique aux relations. Plutôt que de définir une entité qui impose une norme à toutes les instances, nous créons des nœuds et des relations qui peuvent disposer de leur propre jeu de propriétés. Les nœuds de la Figure 1 – Réseau peuvent représenter des clients, des commandes, des produits et des promotions. Chacun d'entre eux peut disposer de types de propriétés différents par type de nœud, et également par instance de chaque nœud. Concernant le nœud client, le prénom, le nom de famille, la date de naissance et la date d'achat peuvent en constituer des propriétés. Mais il se peut que nous ne disposions pas de toutes ces propriétés pour tous les clients. Dans une base de données relationnelle, nous pouvons choisir des propriétés communes et en définir certaines sur NULL en cas d'absence de valeur pour une instance donnée. Cela s'applique également aux relations entre les nœuds : il est possible de stocker des propriétés pour chaque relation sans les imposer à toutes. Des propriétés relatives aux relations peuvent également être stockées. C'est là une différence fondamentale par rapport aux bases de données relationnelles, pour lesquelles nous devons convertir une relation en table pour stocker des métadonnées sur cette relation.

Pour une bases de données de graphes, il n'est pas nécessaire de posséder la même structure pour toutes les instances. Comme nous le verrons ultérieurement, cette souplesse est essentielle dans les histoires de données réelles.

Hiérarchies et arborescences

Une hiérarchie est une structure dans laquelle les nœuds sont organisés les uns au-dessus ou en-dessous des autres. Une arborescence est une hiérarchie comptant plusieurs branches (plusieurs nœuds sont liés à un nœud parent.) Bien sûr, dans le monde réel, les hiérarchies pures sont rares, comme nous le verrons plus en détails dans la prochaine section. Les structures de rapports d'organisation et les missions d'encadrement sont de bons exemples de hiérarchies. Dans une hiérarchie d'encadrement classique, un employé transmet un rapport à un seul autre employé. Ce dernier peut se voir remettre des rapports par plusieurs personnes. Il se murmurerait même que c'est la règle dans ce genre de structure.

Vos données de référence se présentent sous la forme d'un graphe :
Êtes-vous prêt ?

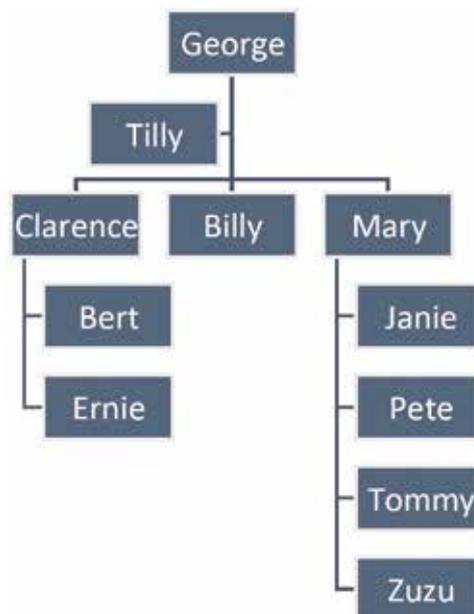


Figure 2 – Hiérarchie employé

Il s'agit d'une situation facile à représenter dans une base de données relationnelle, avec une relation récursive sur une table EMPLOYÉ. Pour une hiérarchie de petite taille, comme l'illustre la Figure 2 - Hiérarchie employé, l'entretien de relations de ce type ne pose aucun problème particulier. Mais dès que l'on modélise un ensemble bien plus vaste, les coûts de maintenance augmentent. Il est souvent nécessaire de trouver des solutions telles que des types de données hiérarchiques spéciaux, ou des colonnes calculées qui effectuent le suivi des niveaux et des pointeurs. Que se passe-t-il lorsqu'un nœud est modifié ? Toutes les relations associées doivent être réinitialisées et recalculées. Si ce nœud appartient à plusieurs types de hiérarchies, il est probable que le nombre de relations à réinitialiser et à recalculer sera important. **Cette histoire de données est tellement complexe qu'il existe des livres spécialisés et des tutoriels sur la conception et la maintenance de ces solutions.** Malgré toutes ces astuces, la maintenance des données et des niveaux de performance d'application reste un problème.

Les hiérarchies du monde réel sont des graphes

Chaque employé représenté dans la Figure 2 - Hiérarchie employé correspond à un nœud. Les lignes de la structure concernant les rapports sont des relations implicites : « Transmet son rapport à ». Toute la structure relative aux rapports peut être parcourue pour trouver tous les employés qui transmettent leur rapport à un employé, et inversement. Mais dans le monde réel, les hiérarchies strictes sont rares. Même dans notre exemple, les employés transmettent souvent leur rapport à plusieurs personnes. Parfois, ces relations existent pour des raisons de transition (observation en situation de travail, partage de tâches, remplacements) et parce qu'il existe de nombreux types de relations « Transmet son rapport à » (responsables, conseillers, gestionnaires de projets, etc.) La Figure 3 - Réseau « Transmet son rapport à » affiche un ensemble de relations plus réaliste et plus riche. Outre le suivi de la transmission directe des rapports aux relations, nous enregistrons les relations administratives, familiales ou encore les relations de protection entre employés. Nous avons de nombreuses relations de transmission de rapport, dont certaines sautent des niveaux dans notre précédent exemple de relations contraintes.

Vos données de référence se présentent sous la forme d'un graphe :
Êtes-vous prêt ?

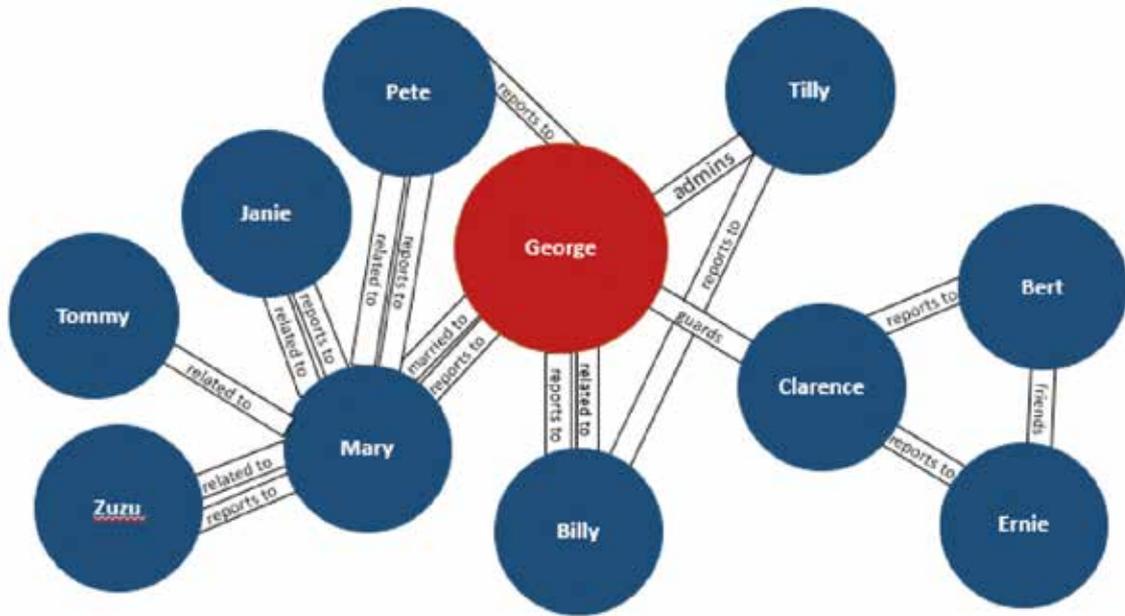


Figure 3 – Réseau « Transmet son rapport à »

Bien qu'on les perçoive souvent comme des hiérarchies, la plupart des hiérarchies d'activité sont en réalité des réseaux. Lorsque vous obtenez une hiérarchie dont la complexité s'apparente à celle de la vie réelle, avec des relations de transmission de rapport multiples et plusieurs types de relations, votre image d'une hiérarchie parfaite et bien ordonnée est réduite à néant. Cela peut se produire dans les organigrammes, les « hiérarchies » de produits, les emplacements et la documentation. Lorsque vous comprenez qu'il est impossible de modéliser votre environnement via une véritable hiérarchie, vous prenez conscience du nombre de graphes qui vous entourent.

Modélisation de données et graphes

La gestion de données traditionnelle prévoit la préparation de deux types de modèles de données : un modèle logique et un modèle physique. Le modèle de données logique décrit les exigences commerciales pour une histoire de données, tandis que le modèle de données physique indique comment les données doivent être conservées dans une base de données. Dans une conception relationnelle, nous appliquons une structure commune à chaque instance d'une entité. Nous avons une entité CLIENT et toutes ces entités partagent le même ensemble de propriétés ou d'attributs. Cela signifie que nous devons découvrir et documenter toutes les propriétés que nous voulons prendre en charge avant de créer la base de données et d'y importer les données.

Dans une bases de données de graphes, le modèle logique correspond au modèle physique. Vous pouvez même vous représenter le modèle de graphe comme un modèle de service. Il est possible de commencer par une modélisation modifiable des nœuds et des relations, puis d'ajouter les propriétés et les étiquettes. À ce stade, nous avons terminé la modélisation des données dont nous avons besoin afin de les saisir dans une bases de données de graphes. Si nous avons des modèles de données logiques traditionnels, nous pouvons même ajouter des propriétés que nous connaissons déjà dans notre modèle de graphe. Comme le modèle logique est le seul modèle, le passage des données aux bases de données est significativement plus rapide et requiert moins de ressources (modélisateurs de données, architectes, administrateur de base de données et développeurs) que la conception de solutions pour données de référence relationnelles.

Nous pouvons également étiqueter chaque instance de façon à pouvoir lancer des requêtes sur nos données en fonction du rôle qu'elles jouent. Par exemple, nous pouvons vouloir lancer certaines requêtes sur les clients de la société uniquement.

La Figure 4 - Rôles des employés, activités, compétences, diplômes, graphe des équipes montre que les nœuds et les relations disposent de propriétés variées. Voilà notre modèle de données, tel qu'il serait dessiné sur un tableau blanc.

Vos données de référence se présentent sous la forme d'un graphe :
Êtes-vous prêt ?

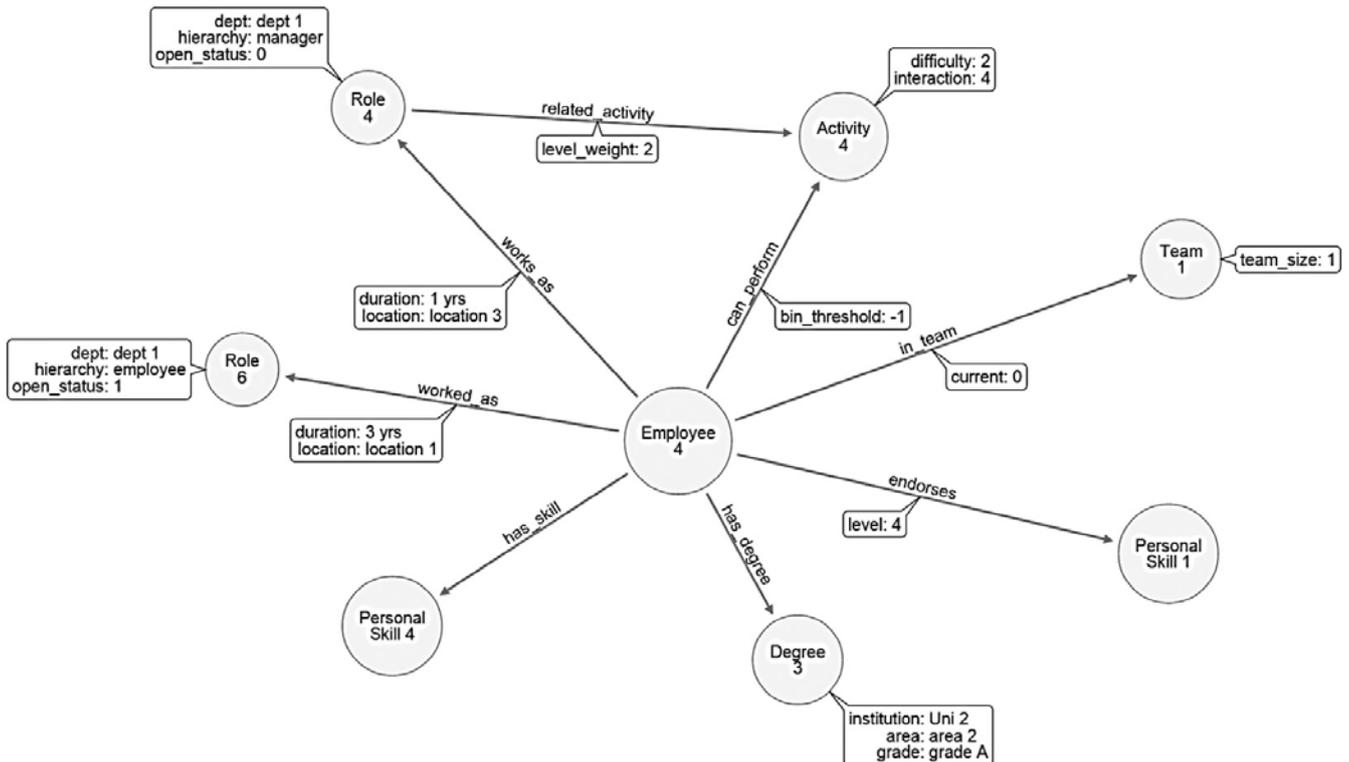


Figure 4 - Rôles des employés, activités, compétences, diplômes, graphe des équipes

Les graphes dans le monde réel

De nombreuses sociétés utilisent des bases de données de graphes pour insister sur les relations de données au niveau de leurs données de référence, et ce dans le but d'être plus compétitives.

Transformation de l'offre dans une entreprise de communication

Une société dans le secteur de l'information et des communications, qui offre des services de téléphonie mobile, de télévision et des services numériques, a revalorisé ses expériences client grâce à l'intégration de Neo4j dans la production de ses offres commerciales. Leur ancien processus d'offre faisait intervenir des vendeurs, et la vente de téléphones mobiles et d'accessoires, les promotions et les plans proposés aux clients entrant dans leur espace de vente nécessitaient l'utilisation de systèmes multiples. La technologie utilisée a été optimisée pour maintenir leurs données produit, et non pour fournir un service client performant, ni aider les gestionnaires de produit à concevoir des lots de produits efficaces. Ils ont essayé d'établir des relations fixes entre les produits et les éléments de services, mais ont ensuite dû payer le prix de l'ajustement et de la redéfinition de ces relations lorsqu'ils ont voulu ajouter de nouvelles propriétés. Ils ont réalisé qu'ils avaient besoin d'une solution plus souple et réactive pour tenir compte des constantes évolutions de leurs articles et des promotions.

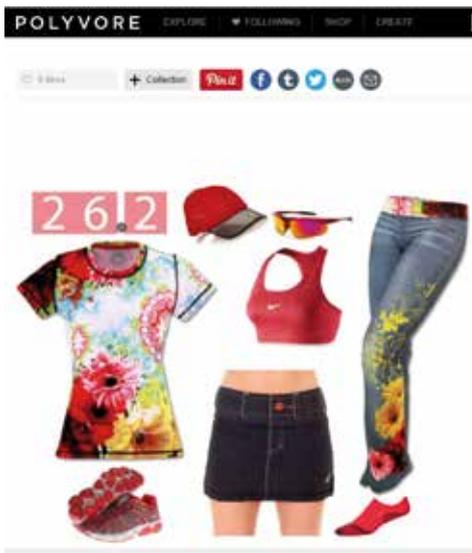
Aujourd'hui, ils utilisent Neo4j, qui permet à leurs clients de sélectionner des groupes d'articles que les vendeurs peuvent rapidement leur préparer. En mettant à profit la souplesse de Neo4j et en insistant sur les relations entre les produits et promotions, ils ont pu réduire le temps de création de lots de produits d'un mois à 10 jours. En magasin, ils peuvent aider leurs clients à choisir les articles qu'ils souhaitent et leur vendre d'autres services ou articles pour leur proposer des lots efficaces. L'utilisation des nombreux types de relations entre les données de référence relatives aux produits et aux clients est cruciale pour l'amélioration du service client et du passage de commande, ainsi que pour réduire le coût des ventes et significativement diminuer le temps d'attente en magasin.

Cette société héberge un million de nœuds et des millions de relations entre ces nœuds, qui continuent à augmenter à mesure que le volume et le nombre d'utilisateurs augmentent. L'utilisation de Neo4j pour les produits et services de télévision câblée et numérique est déjà envisagée afin de bénéficier des mêmes avantages compétitifs au niveau de leurs autres secteurs d'activité.

Exploitation de combinaisons de produits de plusieurs fournisseurs chez Polyvore

POLYVORE

Polyvore permet aux clients de mettre en place des ensembles de vêtements et de produits dans un format hautement graphique et interactif. Un client (ou styliste, selon la terminologie Polyvore) peut choisir des produits, leurs métadonnées et leur prix depuis n'importe quel site Web à l'aide d'un outil de découpage, puis les mettre en valeur avec des graphismes, du texte et des photos personnelles. Le styliste peut ensuite composer des ensembles d'ensembles, appelés « collections », afin que les relations entre les articles soient encore plus nombreuses.



Les utilisateurs de Polyvore composent ces ensembles par millions en une année, multipliant ainsi les intentions d'achat et gagnant en influence auprès des autres clients. Il existe aujourd'hui des millions d'articles au détail (unité de gestion des stocks, UGS) dans cette base de données dont le contenu est généré par les utilisateurs. Les utilisateurs créent ces collections d'articles en fonction de leur goût et de la mouvance.

Polyvore peut demander quels articles « sont assortis » et déduire des réponses les autres produits qui peuvent générer de l'intérêt pour d'autres articles et les vendre. Les utilisateurs peuvent également étiqueter chaque produit avec des caractéristiques pertinentes. Ces étiquettes deviennent des relations pour les données produit, ce qui enrichit davantage le jeu de données de façon participative.

Les marques peuvent également intégrer la communauté Polyvore en plaçant leurs articles dans la base de données dédiée. Elles peuvent par la suite étudier les tendances et la popularité de leurs articles via les commentaires des membres de la communauté et le nombre de « J'aime » sur leurs produits par rapport aux autres.

Ce cas d'utilisation des données produits, des métadonnées et des étiquettes attribuées par les stylistes repose entièrement sur les relations entre ces concepts. Un catalogue produit traditionnel met l'accent sur les produits et leurs propriétés. Mais, dans le cas de Polyvore, l'atout principal vient des relations de données que créent les clients. Les données en elles-mêmes sont presque moins importantes que les relations qui les unissent. Avec l'utilisation de la puissance du graphe dans Neo4j, Polyvore est capable d'identifier les tendances, mais aussi les facteurs qui influencent ces tendances, depuis les prix, styles, couleurs et tailles, jusqu'à n'importe quelle étiquette correspondant aux articles. Les marques ont une meilleure vision de la façon dont leurs produits sont recherchés et perçus par les clients les plus influents du e-commerce.

Pourquoi choisir une base de données de graphes ?

Maintenant que nous avons vu plusieurs utilisations réelles d'une bases de données de graphes, intéressons-nous à son utilisation la plus pertinente. Dans de nombreux cas, le potentiel des bases de données de graphes est clair : médias sociaux, réseaux d'amis d'amis, réseaux de technologie à proprement parler, etc. Ici, les relations entre les utilisateurs et les objets sont clairement des structures de données clé (que nous définissons même comme des réseaux). Mais il existe beaucoup de cas d'utilisation professionnelle où la valeur des relations de données est tout aussi grande, même s'il ne s'agit pas forcément de réseaux.

Vos données de référence se présentent sous la forme d'un graphe :
Êtes-vous prêt ?

À la découverte des nouvelles corrélations entre données

Imaginez que l'on mélange les données d'achat en magasin avec les données tirées de l'expérience d'achat en ligne d'un client :

- Nombre de visites sur le magasin en ligne
- Articles vus, sur lesquels les internautes ont cliqué, ceux qui ont été évalués, ajoutés à la liste d'intention d'achat ou partagés sur les médias sociaux
- Articles ajoutés au panier ainsi que les clics qui ont précédé cet ajout

Ces relations entre les données contiennent de précieuses informations qu'un concurrent voudra découvrir. S'il est possible de les gérer et de les interroger dans une base de données relationnelle, cela risque de s'avérer long et coûteux. De plus, plus le volume de données augmente, plus les coûts d'obtention de résultats en termes de délai de traitement seront importants. Plus on rassemble de données, moins l'obtention de résultats en temps réels sera possible. Penchez-vous sur les questions suivantes susceptibles de se poser par rapport à ces données :

- Quels types d'articles sont achetés ensemble par une catégorie de clients ? Et en considérant une échelle de temps plus longue ?
- Quels sont les membres d'un foyer qui ont tendance à faire quels types d'achat et quels jours de la semaine ? En quoi la distance par rapport au magasin entre-t-elle en jeu ?

Déduplication des données

Il est possible de dupliquer les données de référence dans le cadre de l'acquisition des données d'un tiers ou de la fusion de sociétés. Le recours aux services de prestataires peut également justifier la duplication des données de référence. Les bases de données de graphes peuvent fournir des méthodes plus rapides et plus faciles pour vérifier si « Karen Lopez » à Toronto est le même client que « Karena Lopez », à Scarborough. Il est possible de le savoir en mettant à profit les avantages des requêtes de graphe qui peuvent analyser toutes les relations de données sur plusieurs instances pour déterminer avec certitude que chacun des enregistrements correspond bien à la même personne. Si l'on comprend que deux Karen Lopez, dont les adresses sont différentes, font réparer le même véhicule, ou qu'elles se rendent au même endroit en même temps, cela nous permet de déduire les schémas et les relations entre les données. Nous pouvons utiliser ces relations de données pour identifier les données de référence pour lesquelles il est possible que nous n'ayons pas encore d'identifiant client commun.

Impact et analyse de scénarios

En plus de la gestion des données de référence, les utilisateurs professionnels doivent planifier les modifications à y apporter. Par exemple, les données produites doivent souvent être entièrement repensées et repositionnées. Les responsables du marketing et des produits veulent pouvoir répondre aux questions suivantes :

- Si je change les règles d'utilisation d'une promotion donnée sur certains produits, quel en sera l'impact ?
- Si je modifie le lot de cet ensemble de produits et de services, quelles conséquences cela aura-t-il sur les commandes moyennes ?
- Si l'on modifie ce lot, combien de clients en seront affectés et quelles offres peut-on leur faire en échange ?
- En prenant en compte les données démographiques client d'un nouveau territoire de vente, quels produits et quelles promotions doit-on mettre en avant ?

Toutes ces questions mettent l'accent sur les relations de données entre plusieurs ensembles de données de référence.

Vue à 360 degrés des données de référence

La conception et la construction d'une vue CLIENT ou PRODUIT ne constitue que la première étape. Mais ce n'est pas avec des réserves de données que l'on prend des décisions commerciales. Dans la plupart des cas, nos clients sont des individus qui achètent ou qui souhaitent acheter des produits. Nous sélectionnons des produits parce que nous nous attendons à ce que nos clients les achètent.

Comme nous avons pu le voir avec les histoires de données mentionnées plus haut, les données de référence représentées sous forme de graphe permettent de répondre à ces questions plus facilement, avec plus de souplesse, et sont plus rapides à traiter.

Vos données de référence se présentent sous la forme d'un graphe :
Êtes-vous prêt ?

Architectures modernes des données hybrides

L'architecture des données de référence modernes ne dépend plus d'une seule technologie. De fait, les sociétés les plus concurrentielles ont compris qu'utiliser l'outil adéquat pour réaliser une tâche donnée permet d'obtenir des réponses plus rapidement et donne un meilleur aperçu de la situation.

Présentation d'une architecture moderne des données de référence

Dans la Figure 5 - Données modernes, vous pouvez voir que certaines histoires de données sont mieux modélisées par des bases de données relationnelles. Il existe également des histoires de données qui optimisent plusieurs technologies NoSQL, dont les bases de données de graphes. En fait, un rapport Gartner de janvier 2015 indique que les directeurs des systèmes d'information et les professionnels des données ont besoin de plusieurs technologies de base de données :

Avec le développement des données persistantes multilingues, les directeurs des systèmes d'information et les responsables de la gestion de l'information doivent évaluer de nouvelles approches concernant la cohérence des données, l'optimisation de l'accès aux données et l'intégration des données sur plusieurs types de structures persistantes.¹

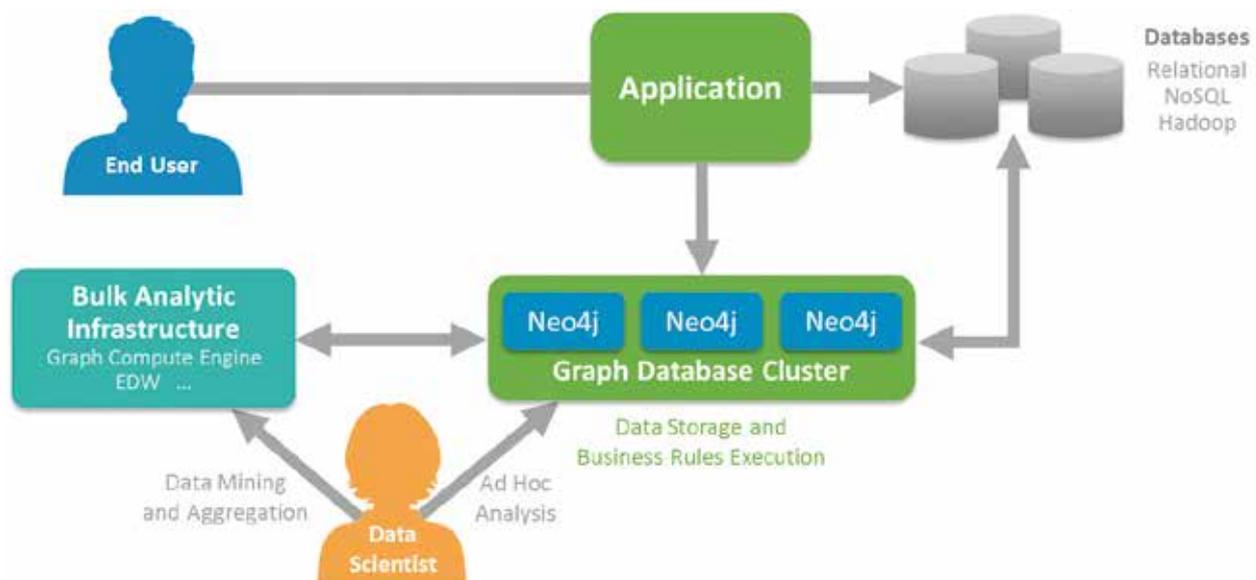


Figure 5 – Architecture moderne des données
(Neo4j. Utilisé avec autorisation)

Ce sont des sociétés qui ne cherchent pas uniquement à rassembler des données, mais à les utiliser pour favoriser la prise de décision en temps réel. Cela implique l'exploitation de plusieurs technologies qui fournissent la meilleure solution pour interroger nos données. Les bases de données relationnelles ne disparaîtront pas pour autant. Nous avons toujours de très bonnes raisons de les utiliser pour les banques de données très structurées et gérées à des fins commerciales. Ces histoires requièrent les types de contraintes et « similitudes » que les bases de données relationnelles ont été conçues pour fournir. Nous avons besoin de ces technologies pour garantir l'intégrité de nos données, pour nous assurer que nous rassemblons toutes les données dont nous avons besoin pour mener à bien une transaction et pour fournir un jeu de données stable et hautement prévisible qui s'avère utile dans nos processus commerciaux. Toutefois, les points forts d'un système relationnel peuvent nuire à la gestion et à l'analyse des données. C'est pourquoi certaines banques de données exigent moins de structure, de contraintes et de cohérences entre plusieurs instances de données. De fait, c'est l'une des raisons principales de la mise en place d'une solution NoSQL.

*Vos données de référence se présentent sous la forme d'un graphe :
Êtes-vous prêt ?*

Performances

Neo4j anticipe l'évolutivité des données. C'est pourquoi les solutions de gestion des données de référence s'adaptent à l'activité commerciale. Comme il s'agit d'une base de données spécifiquement conçue pour stocker et traiter des requêtes de graphe, l'exploitation des relations de données est considérablement plus rapide qu'avec les solutions SQL ou NoSQL.

Souplesse

Une bases de données de graphes permet de maintenir les données sans attributs ni propriétés normatifs applicables à chaque nœud, ou à chaque relation entre les nœuds. Cela constitue un avantage pour nous lorsque nous collectons des données depuis plusieurs sources. Par exemple, nous pouvons avoir des données clients provenant de nos magasins en ligne et de nos magasins physiques, ainsi que des données clients obtenues au moyen d'une liste de diffusion. Il est probable que ces différentes sources partagent certaines données et possèdent des attributs différents, et que certaines sources soient plus fournies que d'autres. Si nous essayions d'importer ces données dans une base de données relationnelle, nous serions forcés de désactiver les contraintes relatives à la qualité des données ou de nettoyer et de rejeter certaines données pour répondre aux règles d'intégrité. Or, nous souhaitons procéder autrement : nous voulons regrouper ces données en une seule base de données que nous pourrions exploiter telle quelle.

Nous avons également la possibilité d'ajouter de nouvelles données, à la volée, parce que nous n'essayons pas de forcer tous les nœuds ou toutes les relations à contenir le même type de métadonnées. Nous pouvons importer les données que nous voulons, les relier aux données existantes (ou pas) et continuer à les interroger. Puisque le modèle logique est le modèle physique, nous n'avons pas à réorganiser l'architecture de notre solution à chaque fois que nous voulons obtenir de nouvelles informations sur nos données de référence.

Cette gestion souple et réaliste des données et des relations de données permettent aux sociétés de poser plus de questions et d'obtenir plus de réponses et plus rapidement qu'avec l'acquisition de données traditionnelle. La possibilité d'ajouter plus de relations de données en temps réel permet aux entreprises de constamment valoriser les histoires de leurs données.

Récapitulatif

Selon nous, l'architecture hybride et multilangage représente l'avenir de la gestion des données de référence pour les organisations compétitives. Nous savons que la valeur de la gestion des données de référence ne repose pas uniquement sur les données elles-mêmes, mais plutôt sur les relations entre ces dernières. Et nous avons vu à quel point il peut être difficile de développer des systèmes rapides et souples pour parcourir ces relations. Les systèmes de relations traditionnels présentent leurs avantages. Toutefois, l'exploration d'un réseau de données de référence est onéreuse, lente et moins réactive au marché.

Les bases de données de graphes telles que Neo4j sont conçues du début à la fin pour servir les histoires de données modélisées par des graphes (ces fameuses histoires que vos données veulent vous conter). Le fait que le modèle de données logique soit le modèle physique signifie que les professionnels des données peuvent apporter des réponses aux questions posées par les relations de données de manière plus souple et plus rapide que jamais.

Vos données de référence se présentent sous la forme d'un graphe :
Êtes-vous prêt ?

Et après ?

Vous venez d'apprendre que les bases de données de graphes jouent un rôle dans l'architecture moderne des données d'entreprise. Qu'allez-vous faire ?

1. Lire Graph Databases (<http://graphdatabases.com>) pour mieux connaître les enjeux théoriques et pratiques de l'application des technologies graphiques.
2. Suivre le cours en ligne sur les bases de données de graphes <http://neo4j.com/graphacademy/online-course>
3. Parler aux principaux utilisateurs professionnels des données de votre entreprise à propos des types de questions qu'ils aimeraient poser à leurs données, mais qu'ils n'ont pas pu interroger pour des questions de coûts ou de limitations technologiques.
4. En savoir plus à propos du passage de SGBDR aux concepts et aux outils de bases de données de graphes <http://neo4j.com/developer/graph-db-vs-rdbms>
5. Aider ces utilisateurs professionnels à comprendre que les relations entre leurs données ne sont sans doute pas mises en évidence dans les structures de données qu'ils utilisent.
6. Télécharger Neo4j <http://neo4j.com> ou utiliser une image machine Amazon (AMI) <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AMIs.html> pour construire un environnement dans le cloud.
7. Découvrir de nouvelles idées à propos de vos données de référence qui donneront à votre société un avantage concurrentiel reposant sur les relations entre les données dont vous disposez déjà.

Un mot sur l'auteur

Karen López est chef de projet principal et architecte chez InfoAdvisors. Elle a passé plus de 20 ans à conseiller les sociétés pour la mise en place de programmes multi-projets de grande ampleur.

InfoAdvisors est une société de conseil en gestion des données, dont le siège social se trouve à Toronto. Nous sommes spécialisés dans l'application pratique de la gestion des données. Notre philosophie consiste à évaluer les coûts, les bénéfices et les risques de toute technique afin de répondre aux besoins précis des sociétés qui font partie de nos clients.

Nous voulons que vous appréciiez vos données.

Retrouvez-nous sur datamodel.com

À propos de Neo Technology

Neo Technology est le créateur de Neo4j, leader mondial des bases de données de graphes, qui propulse les relations entre les données au premier plan. Depuis la recommandation de produits et de services personnalisés par les entreprises, aux sites qui offrent des fonctionnalités de réseaux sociaux, en passant par le diagnostic des problèmes réseau par les opérateurs de services de télécommunications, la réinvention des modèles d'accès, d'identité et de données de référence, les entreprises adoptent les bases de données de graphes pour modéliser, stocker et interroger aussi bien les données que les relations existant entre elles. De grandes entreprises telles que Walmart, eBay, UBS, Nomura, Cisco, HP, Telenor, SFR, Meetic et Glowbl, ainsi que des start-ups telles que CrunchBase, Medium, Polyvore et Zephyr Health utilisent Neo4j pour tirer parti des relations entre les données.

Neo Technology est implantée en Suède, au Royaume-Uni, en Allemagne, en France ainsi qu'en Malaisie, et son siège social se trouve à San Mateo, en Californie. Pour plus d'informations, visitez le site Neo4j.com.