

TP NoSQL

Objectif :

Modéliser, implémenter et tester en volumétrie un service d'analyse de comportement d'achat d'utilisateurs regroupés dans un réseau social. Cette implémentation et ces tests seront effectués avec un SGBDR traditionnel et une base NoSQL afin de comparer les avantages, inconvénients et performances de chaque solution.

Les tests devront pouvoir être effectués par un utilisateur sans intervention dans le code donc il faut également développer un logiciel (Web ou client lourd au choix) permettant de lancer des requêtes sur les 2 bases avec mesure/affichage des temps de réponse.

Cahier des charges :

Les utilisateurs sont regroupés au sein d'un réseau social leur permettant d'avoir des cercles de followers. Le lien de « *follows* » devra être orienté. En termes de volumétrie pour cette phase de test, on peut envisager de créer 1M utilisateurs. Chaque utilisateur pourrait avoir environ 0 – 20 followers directs.

Attention : sur plusieurs niveaux, un utilisateur peut être son propre follower ! Il faut prendre en compte ce point pour éviter, lors des recherches, de doubler les résultats.

Concernant les achats, la base pourrait contenir 10 000 références de produits. Pour les achats, chaque utilisateur pourrait avoir commandé entre 0 et 5 produits parmi ces références.

Méthodologie :

L'alimentation des données se fera par des processus automatisés de génération aléatoire : il ne sera donc pas question ici de juger de la pertinence des résultats obtenus aux requêtes de recherche, mais uniquement du bon fonctionnement des requêtes, et des performances pour leur exécution.

NOTE : il sera peut-être nécessaire d'ajouter manuellement quelques données afin de forcer des réponses pour certaines requêtes si la génération aléatoire ne donne pas entière satisfaction.

L'exploitation / analyse (non exhaustive) des données de la base :

- Obtenir la liste et le nombre des produits commandés par les cercles de followers d'un individu (niveau 1, ..., niveau n) → cette requête permet d'observer le rôle d'influenceur d'un individu au sein du réseau social pour le déclenchement d'achats
- Même requête mais avec spécification d'un produit particulier → cette requête permet d'observer le rôle d'influenceur d'un individu suite à un « *post* » mentionnant un article spécifique
- Pour une référence de produit donné, obtenir le nombre de personnes l'ayant commandé dans un cercle de followers « *orienté* » de niveau n (à effectuer sur plusieurs niveaux : 0, 1, 2 ...) → permet de rechercher les produits « viraux », c'est-à-dire ceux qui se vendent le plus au sein de groupes de followers par opposition aux achats isolés pour lesquels le groupe social n'a pas d'impact
- ...

REMARQUE : si ce sont les performances en recherche/analyse qui sont primordiales pour ce sujet, les temps d'injection dans les bases sont également des données importantes à mesurer et tracer dans le rapport final.

Travail attendu : un rapport comportant les travaux effectués (modèles de données / spécifications & conception du logiciel / requêtes / résultats), les performances obtenues pour chaque système testé (SGBDR **et** NoSQL) à la fois pour les temps d'injection et les temps de recherche, l'analyse de ces résultats et les conclusions qui en découlent sera fourni à l'issue du TP.

Le logiciel permettant à la fois de lancer des injections de données et d'effectuer les différentes recherches mentionnées ci-dessus devra également être développé en respectant les bonnes pratiques de codage, notamment une couche DAL (Data Abstract Layer) permettant d'avoir une IHM unique avec 2 connecteurs vers les 2 types de base, le choix de la base se faisant par l'utilisateur dans l'IHM. Le logiciel doit offrir un maximum de flexibilité (par exemple : pour les injections, le nombre/type d'informations à injecter avec éventuels paramètres d'optimisation, pour les recherches « *en profondeur* », pouvoir spécifier le niveau de profondeur et générer les requêtes en conséquence, ...)