**Mégadonnées – Projet final**

**Données ciblées**

|  |
| --- |
| Liste des produits ciblés sur le site [www.marchevicto.com](http://www.marchevicto.com)  : (# item, produit, producteur, catégorie, etc.) |

**Compréhension du modèle**

|  |
| --- |
| Schéma des tables et champs principaux de la base de données (ceux qui vous permettront de mettre en œuvre les 2 stratégies) : |

**Stratégie #1 (proposée par votre enseignante) :**

Choisissez un produit de votre choix, présent dans la base de données Mon Marché.

En MySQL, trouvez une requête vous permettant de recueillir tous les autres produits présents dans des commandes contenant ce produit. Ajoutez un ou 2 facteurs de votre choix (période de l’année, provenance, fournisseur, etc.)

Trouvez l'occurrence de chacun de ses produits et triez en ordre décroissant d'occurrence (le produit le plus présent apparaissant dans le haut)

Tartinade de Sureau 214035

|  |
| --- |
| Votre requête #1 :  **SELECT** p\_article.id, p\_article.nom, p\_article.gluten, **COUNT**(\*) **AS** occurrence  **FROM** p\_article  **INNER** **JOIN** p\_proto\_article **AS** ppa1 **ON** p\_article.id = ppa1.article\_id  **INNER** **JOIN** p\_commande **ON** ppa1.commande\_id = p\_commande.id  **INNER** **JOIN** p\_proto\_article **AS** ppa2 **ON** p\_commande.id = ppa2.commande\_id  **WHERE** ppa2.article\_id = 214035 **AND** ppa1.article\_id != 214035 **and** p\_article.gluten =1  **GROUP** **BY** p\_article.nom, p\_article.id  **ORDER** **BY** occurrence **DESC**; |

|  |
| --- |
| Tableau des résultats obtenus (10 premiers résultats de sortie de la requête SQL) : |

Temps d’exécution (calculez 10 fois et faire la moyenne) :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. 55 | 1. 55 | 1. 58 | 1. 57 | 1. 57 |
| 1. 2 | 1. 56 | 1. 2 | 1. 2 | 1. 4 |
| **Temps moyen : 34.8** | | | | |

Peaufinez votre requête (réduisez, si possible, le nombre de champs et le nombre de relations, révisez les critères utilisés, etc.) et comparez les temps d'exécution.

|  |
| --- |
| Votre requête #2 :  **SELECT** p\_article.nom, **COUNT**(\*) **AS** occurrence **FROM** p\_article  **INNER** **JOIN** p\_proto\_article **ON** p\_article.id = p\_proto\_article.article\_id  **INNER** **JOIN** p\_commande **ON** p\_proto\_article.commande\_id = p\_commande.id  **WHERE** p\_proto\_article.article\_id != 214035 **AND** p\_commande.id **IN** (**SELECT** commande\_id **FROM** p\_proto\_article **WHERE** article\_id = 214035)  **GROUP** **BY** p\_article.nom  **ORDER** **BY** occurrence **DESC**; |

Temps d’exécution (calculez 10 fois et faire la moyenne) :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. 95 | 1. 62 | 1. 3 | 1. 55 | 1. 54 |
| 1. 55 | 1. 55 | 1. 54 | 1. 2 | 1. 55 |
| **Temps moyen : 49** | | | | |

Modifiez votre requête pour permettre de soumettre un sous-ensemble de 2 produits.  
(Si je choisis les produits A et B, j’aurai la liste des autres produits présents dans des commandes contenant à la fois A et B). Appliquez les mêmes facteurs que ceux établis au départ.

|  |
| --- |
| Votre requête #3 :  **SELECT** p\_article.nom, **COUNT**(\*) **AS** occurrence  **FROM** p\_article  **INNER** **JOIN** p\_proto\_article **ON** p\_article.id = p\_proto\_article.article\_id  **INNER** **JOIN** p\_commande **ON** p\_proto\_article.commande\_id = p\_commande.id  **WHERE** p\_proto\_article.article\_id != 214035  **AND** p\_commande.id **IN** (  **SELECT** commande\_id **FROM** p\_proto\_article **WHERE** article\_id = 214035)  **AND** p\_proto\_article.commande\_id **IN** (  **SELECT** ppa2.commande\_id  **FROM** p\_proto\_article ppa2  **WHERE** ppa2.article\_id = 323133)  **AND** p\_proto\_article.article\_id **NOT** **IN** (214035, 323133)  **GROUP** **BY** p\_article.nom  **ORDER** **BY** occurrence **DESC**; |

Temps d’exécution (calculez 10 fois et faire la moyenne) :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. 65 | 1. 4 | 1. 55 | 1. 2 | 1. 2 |
| 1. 68 | 1. 56 | 1. 68 | 1. 54 | 1. 57 |
| **Temps moyen : 43,1** | | | | |

Modifiez votre requête pour permettre de soumettre un sous-ensemble de 3 produits.  
(Si je choisis les produits A, B et C, j’aurai la liste des autres produits présents dans des commandes contenant à la fois A, B et C). Appliquez les mêmes facteurs que ceux établis au départ.

|  |
| --- |
| Votre requête #4 :  **SELECT** p\_article.id, p\_article.nom, **COUNT**(\*) **AS** occurrence  **FROM** p\_article  **INNER** **JOIN** p\_proto\_article **ON** p\_article.id = p\_proto\_article.article\_id  **INNER** **JOIN** p\_commande **ON** p\_proto\_article.commande\_id = p\_commande.id  **WHERE** p\_proto\_article.article\_id != 214035  **AND** p\_commande.id **IN** (  **SELECT** commande\_id **FROM** p\_proto\_article **WHERE** article\_id = 214035)  **AND** p\_proto\_article.commande\_id **IN** (  **SELECT** ppa2.commande\_id **FROM** p\_proto\_article ppa2 **WHERE** ppa2.article\_id = 323133)  **AND** p\_proto\_article.commande\_id **IN** (  **SELECT** ppa3.commande\_id **FROM** p\_proto\_article ppa3 **WHERE** ppa3.article\_id = 518077)  **AND** p\_proto\_article.article\_id **NOT** **IN** (214035, 323133, 518077)  **GROUP** **BY** p\_article.nom, p\_article.id  **ORDER** **BY** occurrence **DESC**; |

Temps d’exécution (calculez 10 fois et faire la moyenne) :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. 56 | 1. 58 | 1. 2 | 1. 54 | 1. 2 |
| 1. 54 | 1. 56 | 1. 57 | 1. 54 | 1. 56 |
| **Temps moyen : 44,9** | | | | |

Ajoutez des Index (<https://dev.mysql.com/doc/refman/8.0/en/mysql-indexes.html>) et validez leurs impacts sur les performances de vos requêtes.

|  |
| --- |
| Vos index:  **CREATE** **INDEX** idx\_orders\_customer\_id **ON** p\_commande(id);  **CREATE** **INDEX** idx\_proto\_article\_id **ON** p\_proto\_article(id);  **CREATE** **INDEX** idx\_article\_id **ON** p\_article(id); |

Temps d’exécution (calculez 10 fois et faire la moyenne) :

Requête #1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. 59 | 1. 55 | 1. 62 | 1. 1 | 1. 1 |
| 1. 56 | 1. 2 | 1. 54 | 1. 1 | 1. 4 |
| **Temps moyen : 29.5** | | | | |

Requête #2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. 1 | 1. 1 | 1. 57 | 1. 1 | 1. 0 |
| 1. 53 | 1. 2 | 1. 54 | 1. 1 | 1. 54 |
| **Temps moyen : 22.4** | | | | |

Requête #3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. 66 | 1. 56 | 1. 59 | 1. 59 | 1. 57 |
| 1. 1 | 1. 61 | 1. 56 | 1. 59 | 1. 2 |
| **Temps moyen : 47.6** | | | | |

Requête #4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. 1 | 1. 56 | 1. 58 | 1. 56 | 1. 61 |
| 1. 56 | 1. 62 | 1. 2 | 1. 65 | 1. 56 |
| **Temps moyen : 47.3** | | | | |

**Analyse des résultats de la stratégie #1 :**

Est-ce que la solution proposée permet de faire des croisements entre les produits ciblés?

Qu’avez-vous appris en réalisant mettant en œuvre cette stratégie?

Quels sont selon vous, les avantages (2) et inconvénients de la stratégie proposée?

|  |
| --- |
| Votre analyse (environ 250 mots) : |

**Stratégie #2 (mise en œuvre par l’élève, individuellement) dans un SGBD NoSQL**

|  |
| --- |
| Choix et justification de la stratégie sélectionnée : |

|  |
| --- |
| Requêtes d’exportation (joindre les fichiers exportés dans la remise du projet) : |

|  |
| --- |
| Script de conception du modèle et d’importation des données : |

|  |
| --- |
| Requête #1 ayant comme objectif de répondre au besoin du client : |

|  |
| --- |
| Requête #2 ayant comme objectif de répondre au besoin du client : |

|  |
| --- |
| Requête #3 ayant comme objectif de répondre au besoin du client : |

|  |
| --- |
| Laquelle de ces 3 requêtes vous semble la plus pertinente?  Justifiez :  Est-ce possible de l’optimiser en modifiant les opérations appliquées sur la requête (conditions, jointures, etc.)?  Si oui, inscrire ici la requête optimisée :  Si non, justifier : |

Temps d’exécution de la requête sélectionnée (calculez 10 fois et faire la moyenne) :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
| **Temps moyen :** | | | | |

Ajoutez des index et validez leurs impacts sur les performances de vos requêtes.

|  |
| --- |
| Vos index: |

Temps d’exécution de la requête sélectionnée (calculez 10 fois et faire la moyenne) :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
| **Temps moyen :** | | | | |

**Analyse des résultats de la stratégie #2 :**

Est-ce que la solution proposée permet de faire des croisements entre les produits ciblés?

Qu’avez-vous appris en réalisant mettant en œuvre cette stratégie?

Quels sont selon vous, les avantages (2) et inconvénients de la stratégie proposée?

|  |
| --- |
| Votre analyse (environ 250 mots) : |