

Introduction NOSQL

Table des matières

I. Qu'est-ce qu'une base de données NoSQL ?	3
II. Exercice : Quiz	5
III. Comment configurer une base de données NoSQL ?	6
IV. Exercice : Quiz	10
V. Essentiel	11
VI. Auto-évaluation	12
A. Exercice	12
B. Test	12
Solutions des exercices	12

I. Qu'est-ce qu'une base de données NoSQL ?

Durée : 1 h

Environnement de travail : PC

Contexte

Inventé par Carl Strozzi en 1998, le concept et le terme NoSQL servait à cette époque à désigner une base de données relationnelle open source et légère. Quelques années plus tard, NoSQL a été adopté et vulgarisé par les grandes firmes comme Amazon, Google, Facebook. Ces derniers étaient confrontés à des données aux volumes très importants. La lenteur des bases de données relationnelles étant l'une des raisons incitant les grandes firmes à se tourner vers ce concept.

Les géants de la technologie ont préféré répartir la charge sur plusieurs serveurs hôtes plutôt que de mettre à jour leurs matériels informatiques pour étendre les performances des Relational DataBase Management System (RDBMS). Cette méthode est appelée « *scaling out* ». Non relationnelles, les bases de données NoSQL sont parfaites pour cette méthode.

L'histoire du NoSQL a également été marquée par Amazon Dynamo en 2007. L'année suivante, Facebook rend la base de données non-relationnelle qu'elle utilise en interne open source. Cet outil nommé Cassandra devient alors une référence en matière de databases NoSQL, et donne au concept NoSQL sa popularité.

NoSQL peut être une option de base de données pour développer une application. Une base de données NoSQL peut aussi servir pour tester une application sur un serveur local. C'est pourquoi, il est nécessaire de savoir la créer et la configurer.

Définition **La base de données NoSQL**

Le NoSQL ou Not Only SQL est une catégorie de base de données non-relationnelle. Servant d'alternative au langage SQL, le NoSQL s'appuie sur des documents et des collections plutôt que sur des tables et des lignes. Dans le NoSQL, une collection désigne une agrégation de documents. Ces derniers constituent des objets JSON individuels, présentés sous forme de paires clé-valeur. Dans le NoSQL, une collection peut être composée de groupes d'informations. Les documents d'une collection ne possèdent pas de structure définie. Ils sont également indépendants les uns des autres.

Exemple

Les bases de données non-relationnelles les plus populaires sont :

- MongoDB
- Cassandra
- Couchbase
- Document DB
- Redis
- Neo4j
- HBase

Généralement, les bases de données NoSQL reposent sur des données dénormalisées. Elles prennent en charge des applications utilisant peu de conteneurs ou de tables avec des relations non modélisées par les références au niveau des données. Les types d'applications s'appuyant sur des données très normalisées de peur de rencontrer des anomalies ou une duplication des données ne conviennent pas à NoSQL.

Les bases de données NoSQL possèdent des requêtes très complexes. Elles fonctionnent parfaitement bien avec les requêtes sur une table unique. Toutefois, plus les requêtes augmentent, plus les bases de données relationnelles deviennent le meilleur choix. En effet, les bases de données NoSQL n'offrent pas de jointures complexes. Elles n'offrent également ni de sous-requêtes, ni d'imbrication de requêtes.

Aujourd'hui, de plus en plus d'entreprises optent pour un modèle convergent pouvant combiner des modèles de données non-relationnelles et relationnelles. Opter pour ce type d'approche permet une plus grande flexibilité dans la gestion des données et assure une bonne lecture et une écriture cohérente tout en restant performante.

Attention

Les bases de données NoSQL sont fréquemment utilisées dans :

- Le Big Data
- Les applications web en temps réel

Les développeurs apprécient les bases de données NoSQL. Ces dernières sont pratiques pour le développement agile. En effet, elles peuvent rapidement s'adapter aux besoins. Par ailleurs, les bases de données NoSQL permettent de sauvegarder des données de façon intuitive. Elles réclament peu de transformations lors du stockage des données ou au moment de la récupération d'API de type NoSQL. Les bases de données NoSQL peuvent tirer parti du Cloud. Cela, pour n'offrir aucun temps d'arrêt.

Les bases de données NoSQL peuvent être utilisées dans de nombreuses situations.

Exemple

Situation nécessitant l'utilisation de bases de données NoSQL :

- Pour le stockage d'immenses volumes de données.
- Quand les relations entre les données ne sont pas importantes.
- Lorsque les données rencontrent des changements au fil du temps et sont déstructurées.
- Quand le volume de données s'accroît de manière continue et nécessite un scaling fréquent de la base de données.

Remarque

Les avantages d'opter pour une base de données NoSQL non relationnelle :

- **Une évolutivité** : les bases de données NoSQL peuvent se déployer en ayant recours à du matériel de base. Ainsi, elles peuvent supporter un trafic accru et répondre à la demande sans s'arrêter. Les bases de données NoSQL peuvent gagner en puissance et en performance en se développant. C'est pourquoi, elles sont devenues appréciées les données en évolution.
- **Une flexibilité** : NoSQL permet de stocker des données de façon libre sans schéma rigide. Grâce au NoSQL, les applications peuvent être développées rapidement et devenir plus innovantes. Il permet aux développeurs de se concentrer davantage sur la création de systèmes afin de mieux servir les clients sans rencontrer des difficultés au niveau des schémas.
- **Une grande performance** : une base de données NoSQL possède une architecture évolutive. Cela peut être très utile lorsque le trafic ou le volume de données s'accroît. Cette architecture assure des temps de réponse optimisés et prévisibles. Les bases de données NoSQL sont également capables d'ingérer des données et les

fournir de manière rapide et fiable. Tous ces avantages font qu'elles sont très appréciées dans les applications qui collectent régulièrement des téraoctets de données tout en demandant une expérience utilisateur ultra interactive.

- **Une fonctionnalité optimale** : élaborée pour les magasins de données distribués, les bases de données NoSQL peuvent stocker des données très importantes.
- **Une disponibilité** : les bases de données NoSQL permettent de réduire de façon efficace la latence pour les utilisateurs, quel que soit l'endroit où ils se trouvent. Ses fonctionnalités permettent par ailleurs de diminuer la charge de gestion des bases de données. Ainsi, les utilisateurs peuvent se concentrer sur leur cœur de métier.



Comprendre les bases de données non relationnelles

Exemple

Les bases de données NoSQL :

- Peuvent traiter rapidement tous types de formats de données, à savoir les données structurées, les données semi-structurées et les données non-structurées dans un unique magasin de données.
- Sont un choix judicieux pour le Big Data, le client 360, les applications web en temps réel, les jeux ou les achats en ligne, les réseaux sociaux, l'internet of things, les applications de pub en ligne.

Exercice : Quiz

[solution n°1 p.13]

Question 1

Le NoSQL fait référence à des bases de données relationnelles.

- ☐ Vrai
- ☐ Faux

Question 2

Les bases de données NoSQL sont appréciées dans la réalisation d'applications web en temps réel et du Big Data.

- ☐ Vrai
- ☐ Faux

Question 3

Les bases de données NoSQL peuvent traiter tous types de formats.

- ☐ Vrai
- ☐ Faux

Question 4

Cassandra est une base de données relationnelles.

- ☐ Vrai
- ☐ Faux

Question 5

Les bases de données NoSQL permettent aux entreprises de rester agile et d'innover.

- ☐ Vrai
- ☐ Faux

III. Comment configurer une base de données NoSQL ?

Il est essentiel de différencier le NoSQL du SQL. Ils possèdent tous les deux des cas d'utilisation particuliers.

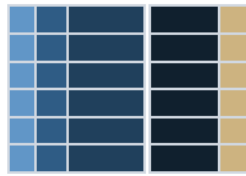
- Les bases de données NoSQL stockent des données qui sont accessibles sous forme d'objets JSON non structurés.
- SQL sauvegarde les informations dans des fichiers plus organisés.
- NoSQL emploie le mappage de document ODM pour relier les collections dans une base de données.
- SQL relie les tables dans un mappage objet relationnel ORM.
- Les schémas NoSQL sont flexibles. Le concept ne requiert aucune jointure harassante.
- SQL possède des schémas rigides. Les requêtes SQL sont souvent pâles comparées à celles du concept NoSQL.
- La majorité des bases de données NoSQL sont des logiciels Open Source. Les internautes peuvent y avoir accès librement. Ils peuvent les télécharger gratuitement via les applications NoSQL. Ils peuvent se familiariser avec et les utiliser en supplément de leur base de données SQL.

Ces bases de données sont les plus privilégiées et ont fait leurs preuves :

- Cassandra : en colonne, cette base de données NoSQL est très utile pour le traitement de jeu et le stockage de jeux de données importants.
- Redis : cette base de données clé-valeur est une base de données en mémoire. Elle sauvegarde les données dans le cache et assure aux utilisateurs une meilleure performance.
- CouchDB : cette base de données documentaire peut être utilisée dans tous les domaines d'application. Elle exploite une API intuitive http/JSON.
- Neo4j : elle permet de représenter les relations entre les divers jeux de données et de reconnaître les modèles à définir dans les graphiques.

Base de données SQL

Relationnelle

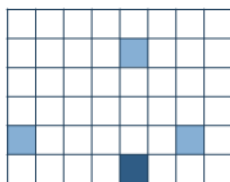


OLAP



Base de données NoSQL

Orienté colonne



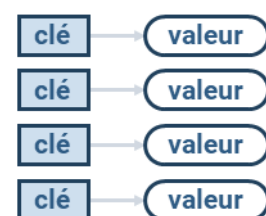
Document



Graphe



Clé-Valeur



Base de données SQL et base de données NoSQL

Exemple

Modèle de développement

- SQL : open source comme MySQL ou logiciel payant comme base de données Oracle,
- NoSQL : open Source ou logiciel payant.

Dimensionnement

- SQL : Vertical
- NoSQL : Horizontal

API :

- SQL : l'appel de données et les requêtes d'enregistrement passent par le SQL,
- NoSQL : les données sont consultées et enregistrées à partir d'API orienté objet.

Schéma :

- SQL : la structure et le type des données sont définis en amont. La base de données doit être ajustée afin de pouvoir enregistrer des informations récentes.
- NoSQL : le schéma est flexible. Il permet d'ajouter de nouveaux jeux de données de façon immédiate. Cette base de données permet d'enregistrer les données structurées et non structurées.

Le stockage de données :

- SQL : toutes les données sont enregistrées dans des lignes de tableau et destinées à des attributs spécifiques.
- Les jeux de données sont enregistrés à l'intérieur des tableaux et sont regroupés par le système au moment de recherches complexes.
- NoSQL : les tableaux ne sont pas nécessaires pour ce type de base de données. La base de données se sert de documents entiers, de colonnes, de graphiques ou de key values.

Méthode

Méthode pour la configuration d'une base de données MongoDB sous Windows :

- Il est nécessaire de télécharger et d'installer certains fichiers pour exécuter MongoDB sur l'ordinateur.
- Le plus important, c'est d'installer le serveur MongoDB.

Pour commencer :

1. Télécharger MongoDB et installer la version communautaire du serveur du même nom.
Pour cela, il est nécessaire de développer le MongoDB Community Server tab une fois sur la page de téléchargement.
2. Cliquer sur **Télécharger** pour avoir une version compatible avec le système d'exploitation utilisé. Cette option se trouve à droite, dans la barre latérale.
3. À la fin du téléchargement, ouvrir le fichier d'installation pour démarrer l'installation du serveur MongoDB sur l'ordinateur.
4. Cliquer sur suivant et accepter le contrat de licence.
5. Cliquer sur suivant.
6. Copier le chemin du fichier dans le **Répertoire de données** et le coller ailleurs. Ce répertoire est modifiable.
7. Remplacer les **données** dans le chemin par un chemin bin
Exemple : C:\Program Files\MongoDB\Server\5.0\data\
Remplacer par :
C:\Program Files\MongoDB\Server\5.0\bin\
8. Cliquer **Suite**. Celui-ci est installé dans le processus d'installation.
9. Sélectionner **Complet**.
10. Attendre l'installation de la boussole MongoDB puis cliquer sur **Suivant**.
11. Cliquer **Installer** pour mettre le serveur sur l'ordinateur.

Méthode

Créer et interroger une table NoSQL

- Étape 1 : procéder à la création d'un NoSQL table,
- Étape 2 : intégrer des données au NoSQL table,
- Étape 3 : interroger le NoSQL Table,
- Étape 4 : annuler un élément existant,,
- Étape 5 : annuler un NoSQL table.

Exemple

Une fois la base de données MongoDB installée, il est possible pour l'utilisateur de créer des applications sur son ordinateur. Si l'utilisateur installe MongoDB Compass, il peut afficher les collections et les bases de données.

Attention

Les bases de données NoSQL possèdent divers types de technologies qui possèdent chacune un domaine d'application particulier.

Il existe 4 bases de données NoSQL :

- **La base de données orientées colonnes**

Similaire aux bases de données relationnelles, ce type de base de données s'adapte à la capacité de s'adapter parfaitement aux ajouts et aux diverses manipulations.

Les colonnes sont dynamiques dans le NoSQL, les lignes peuvent avoir un nombre de colonnes différent sans laisser apparaître la valeur NULL.

Dans ce type de base de données, les colonnes d'une table sont stockées dans un fichier unique avec un nombre de lignes fixes. Ainsi, il n'est pas nécessaire de télécharger toute la table lorsque l'utilisateur désire appeler une ligne de cette table. Le NoSQL offre un gain de performance grâce à la requête sur une ligne directement extraite du Cloud.

- **La base de données orientée clé valeurs**

La donnée est stockée sous forme de couple clé/valeur. La donnée valeur est labellisée avec une clé.

Une clé concerne l'information numérique ou écrite sur une donnée.

La valeur désigne la donnée.

La base de données orientée clé valeur comporte beaucoup de textes et d'informations concernant les profils des responsables de textes. Ces derniers forment un bloc de données. Ici, l'idée, c'est de découper le bloc en petits textes. Ces derniers vont être stockés sur les serveurs.

Les petits blocs doivent être labellisés afin de réaliser correctement un réagencement. Dans cette base de données, chaque morceau contient une valeur. Tous les morceaux sont labellisés par une clé. Le but étant de faire le maximum afin de ne pas oublier un post. La fonction qui partage le bloc est nommée fonction de hachage.

- **La base de données orientées Document**

Les données en clé/valeur sont stockées. La valeur d'un document abrite toutes les informations de la data. Généralement, la valeur est un document JSON. Ce dernier est très apprécié des Data Analysts. La valeur peut également être un document TXT, BSON, YAML ou XML.

- **Les bases de données graph-based**

Elles sont composées de bords et de nœuds. Les bords représentent les relations entre les nœuds tandis que les nœuds symbolisent la valeur de la donnée.

Les bases de données graph based permettent une analyse rapide des relations entre les données.

Dans les réseaux sociaux par exemple, les profils constituent les nœuds et les amitiés représentent les bords.

Un graphe bénéficie d'une bonne lisibilité. Les informations concernant la database et les relations peuvent être facilement comprises.

Les inconvénients des bases de données NoSQL :

Elles ne prennent pas en compte les propriétés ACID qui signifient Atomicité, Cohérence, Isolation, Durabilité dont bénéficient les bases de données relationnelles. Pour une prise en charge ACID, les utilisateurs doivent créer leurs propres codes.

Les bases de données NoSQL ne possèdent pas de règles de standardisation. Les données sont dénormalisées. Elles traitent avec des enregistrements sur les serveurs. Ce qui n'est pas le cas du SQL qui fonctionne par le référencement.

Les utilisateurs préconisent les données normalisées, car elles permettent de prévenir les anomalies et les duplications de données.

Dans le cas d'une transaction financière par exemple, les règles de normalisations sont strictes. L'utilisation des données qui ne sont pas normalisées peut alors entraîner des erreurs au niveau de la transaction. Cela pourrait avoir de graves répercussions pour les utilisateurs.

Le NoSQL ne peut pas manipuler plusieurs tables. Avec ce type de base de données, la réalisation de jointures complexes est impossible.

Il n'existe aucune compatibilité entre NoSQL et SQL. Cela, même si les bases de données NoSQL emploient un langage de requête structuré.

Exemple

Utilisation de bases de données orientées colonnes :

Très utile pour gérer des données contenant des clés et des colonnes. Les bases de données orientées colonnes permettent de traiter convenablement les données proches des tables de bases relationnelles sans avoir à rencontrer des lourdeurs.

Utilisation de bases de données orientées clés valeurs :

Ce type de bases de données est idéal pour la gestion des données séparables par fonction de hachage. Les bases de données orientées clés valeurs ont la capacité de stocker d'importants volumes de données Cloud dans de multiples serveurs. La base de données orientée clés valeurs est très légère et est facilement récupérable grâce à sa clé. La base de données est très rapide et très flexible.

Utilisation de bases de données orientées document :

La base de données permet de classer des documents de manière digitalisée grâce à des serveurs en Cloud. Les bases de données orientées document utilisent une clé de série et une valeur.

Ce type de base de données permet de stocker d'importante quantité de données sur de multiples serveurs. Ces données peuvent ensuite être récupérées via des requêtes qui feront appel aux clés de chaque document. Les bases de données orientées document sont très utiles pour stocker des données sous formes de documents réunissant de nombreuses informations.

Utilisation de bases de données graphiques :

Très adapté aux exigences des réseaux sociaux, ce type de base de données est compatible avec les algorithmes de parcours de graphe. Ces derniers permettent l'exploration de données de façon très efficace.

Exercice : Quiz

[solution n°2 p.13]

Question 1

NoSQL relie les tables dans un système de mappage objet relationnel ORM.

- ☐ Vrai
- ☐ Faux

Question 2

JSON est un NoSQL.

- ☐ Vrai
- ☐ Faux

Question 3

Les schémas NoSQL sont rigides.

- ☐ Vrai
- ☐ Faux

Question 4

La Database Graph-based est une base de données composée de nœud.

- ☐ Vrai
- ☐ Faux

Question 5

Les bases de données orientées document permettent de digitaliser les documents.

- ☐ Vrai
- ☐ Faux

V. Essentiel

La spécificité du NoSQL est le fait qu'il soit non relationnel. Ce type de base de données permet de stocker et d'analyser le Big Data.

Comme le Big Data entre aujourd'hui dans de nombreux domaines, les bases de données relationnelles ne sont plus conformes aux besoins. La prise en charge, le stockage et l'analyse de données requièrent des solutions innovantes.

Dans un NoSQL, les données peuvent être stockées de manière non structurée sans obéir au schéma fixe. De ce fait, les jointures ne sont plus utiles et le scaling est facilité.

Le NoSQL est préconisé pour les besoins importants des Data Stores distribués. De nombreux géants de la technologie à l'instar de Google ou de Twitter ont recours au NoSQL, car ils collectent quotidiennement des terabytes de données concernant les utilisateurs.

Les systèmes NoSQL ou Not Only SQL peuvent s'adapter à une large gamme de technologies permettant de stocker des données structurées, non structurées, semi-structurées ou polymorphiques.

Les raisons de choisir une base de données NoSQL sont diverses :

- Pour le stockage de volumes de données très importants qui peuvent posséder ou non une structure. NoSQL a l'avantage de ne pas limiter les types de données à stocker.
- Pour obtenir le meilleur parti du Cloud Computing.
- Pour obtenir le meilleur parti du stockage.
- Pour optimiser le développement.
- Pour augmenter l'évolutivité horizontale.

VI. Auto-évaluation

A. Exercice

Vous êtes chargé de gérer le stock de produits de votre entreprise. Votre supérieur met à votre disposition différents types de bases de données non relationnelles.

Question 1

[solution n°3 p.15]

Quel type de bases de données non relationnelles serait idéal pour classer vos documents ?

Vous avez à votre disposition une base de données SQL et une base de données NoSQL.

Question 2

[solution n°4 p.15]

Quels sont les avantages des bases de données NoSQL comparés à la base de données SQL ? Quelles bases de données choisir ?

B. Test

Exercice 1 : Quiz

[solution n°5 p.15]

Question 1

Les bases de données relationnelles ne sont pas adaptées au Big Data.

- ☐ Vrai
- ☐ Faux

Question 2

Le NoSQL permet de stocker des données sous une forme non structurée.

- ☐ Vrai
- ☐ Faux

Question 3

Les jointures sont nécessaires dans une base de données NoSQL.

- ☐ Vrai
- ☐ Faux

Question 4

Le NoSQL est préconisé pour les besoins de stockage volumineux.

- ☐ Vrai
- ☐ Faux

Question 5

Le NoSQL ne permet pas le stockage de données polymorphiques.

- ☐ Vrai
- ☐ Faux


Solutions des exercices

Exercice p. 5 Solution n°1**Question 1**

Le NoSQL fait référence à des bases de données relationnelles.

☐ Vrai

☒ Faux


 Le terme NoSQL désigne des types de bases de données non relationnelles. Le stockage des données diffère des tables relationnelles.

Question 2

Les bases de données NoSQL sont appréciées dans la réalisation d'applications web en temps réel et du Big Data.

☒ Vrai

☐ Faux


 Les bases de données NoSQL bénéficient d'une forte disponibilité et d'une grande évolutivité, c'est pourquoi elles sont utiles dans la réalisation d'applications web en temps réel et du Big Data.

Question 3

Les bases de données NoSQL peuvent traiter tous types de formats.

☒ Vrai

☐ Faux


 Elles peuvent traiter des données structurées, semi-structurées et non structurées.

Question 4

Cassandra est une base de données relationnelles.

☐ Vrai

☒ Faux


 Cassandra fait partie des bases de données non-relationnelles.

Question 5

Les bases de données NoSQL permettent aux entreprises de rester agile et d'innover.

☒ Vrai

☐ Faux

 Les entreprises peuvent profiter d'une grande flexibilité et d'une multitude de modèles de données. Ce qui permet aux entreprises de rester compétitives.


Exercice p. 10 Solution n°2

Question 1

NoSQL relie les tables dans un système de mappage objet relationnel ORM.

☐ Vrai

☒ Faux


 NoSQL utilise un mappage de document objet ODM pour joindre les collections à l'intérieur d'une base de données.

Question 2

JSON est un NoSQL.

☒ Vrai

☐ Faux


 Il est utilisé pour certain NoSQL de magasins de données.

Question 3

Les schémas NoSQL sont rigides.

☐ Vrai

☒ Faux


 Ce sont les schémas SQL qui sont rigides, les schémas NoSQL sont flexibles.

Question 4

La Database Graph-based est une base de données composée de nœud.

☒ Vrai

☐ Faux


 Elle est composée de nœuds et de bords.

Question 5

Les bases de données orientées document permettent de digitaliser les documents.

☒ Vrai

☐ Faux

 Ils permettent de digitaliser une bibliothèque de dossiers.

p. 12 Solution n°3

La base de données orientée document est idéale pour stocker et digitaliser des dossiers. S'il s'agit de stock de produit, alors, cette base de données va classer les documents grâce à des clés. Ces dernières comportent, à titre d'exemple, le numéro de série des articles et l'intérieur du dossier, une série d'informations concernant le produit, tout cela de manière digitalisée.

p. 12 Solution n°4


Le choix de la base de données à utiliser dépend avant tout de l'objectif de l'application. Le SQL se sature rapidement lorsqu'il s'agit de l'utiliser avec les Big Data. NoSQL quant à lui présente de nombreux avantages, car il peut lire et traiter des volumes de données très importants, et cela, à grande vitesse. Le NoSQL se démarque également par ses schémas flexibles. Les bases de données NoSQL sont facilement accessibles, elles se présentent sous forme de logiciels Open Source. Elles peuvent être utilisées comme complément d'une base de données SQL.

Exercice p. 12 Solution n°5**Question 1**

Les bases de données relationnelles ne sont pas adaptées au Big Data.

☒ Vrai

☐ Faux


 Depuis l'arrivée du Big Data, les bases de données non relationnelles sont mieux adaptées à stocker d'immenses volumes de données.

Question 2

Le NoSQL permet de stocker des données sous une forme non structurée.

☒ Vrai

☐ Faux


 Les données peuvent être stockées sous forme non structurée, et cela, sans suivre de schéma fixe.

Question 3

Les jointures sont nécessaires dans une base de données NoSQL.

☐ Vrai

☒ Faux

 Les jointures ne sont pas nécessaires dans une base de données NoSQL. Le scaling est également facilité.

Question 4

Le NoSQL est préconisé pour les besoins de stockage volumineux.

☒ Vrai

☐ Faux

Q Le NoSQL est utile pour les stockages de données volumineux comme les Data Stores distribués.

Question 5

Le NoSQL ne permet pas le stockage de données polymorphiques.

☐ Vrai

☒ Faux

Q Il permet le stockage de données de différentes structures, et notamment les données polymorphiques.