Les bases de données NoSQL

L. KERKENI

IUT de Laval - Dpt. MMI

2019-2020

Retour sur la notion de "base de données"

Par définition, une base de données c'est :

1. Un ensemble **structuré** d'informations,
2. stocké de manière **persistante** pour être préservé à long terme.

Indissociable d'un système informatique (le SGBD) qui assure des services

1. D'accès aux fichiers (sur un disque)
2. De recherche
3. de sécurisation,
4. et beaucoup d'autres...

Bases relationnelles

Les **bases relationnelles** sont utilisées dans toutes les applications gérant des informations **structurées** et **régulières**.

- applications de "gestion",
- applications Web type commerce électronique,
- applications mobiles.

Les SGBD relationnels apportent de très nombreuses fonctionnalités : stockage, interrogation SQL, concurrence d'accès, reprise sur panne, droits d'accès, optimisation, environnement de développement, etc.

Problématique 1 : La notion de "document"

Document = unité d'information autonome ou quasi-autonome.

- Peu ou pas de référence à d'autres documents.
- Peu ou pas de structure ; ou une structure très flexible.
- Un contenu souvent à orientation multimédia.

Examples (1) : documents textuels, types documents Web.

Examples (2) : images, documents audios, des vidéos ; pas de structure explicite, production de descripteurs synthétiques pour tenter de les indexer.

Examples (3) : jeux en ligne : objets 3D, actions utilisateur.

Problématique 2 : données à très grande échelle

On atteint facilement des **volumes** de données extrêmement importants.

- les moteurs de recherche qui collectent des **documents** disponibles sur le Web.
- les applications utilisées à l'échelle du Web ; commerce électronique (Amazon) ; réseaux sociaux (Facebook).
- données gérées par les jeux en ligne.

Les collections occupent typiquement des centaines de Gigaoctets, voire des Téraoctets.

Solution

Nouveaux systèmes, dits "NoSQL" pour gérer de vastes collections de documents et passer à l'échelle.

Bases NoSQL

3 aspects qui ont motivé la nouvelle vague de bases NoSQL :

- Le VOLUME des données créées double tous les 2 ans. IDC estime qu'en 2020 le volume atteindra 44 Zettabytes (1 ZB = 1 milliards de terabytes)
- La VARIÉTÉ des types de données créées (Smartphones)
- La VÉLOCITÉ avec laquelle les données changent est également très importante (Internet of Things)

Bases de données orientées documents

Deux langages de représentation : JSON et XML.
Un SGBD adapté :

- basé sur JSON : MongoDB
- basé sur XML : eXist

Représentation de documents textuels

Modèle de données basé sur des **graphes**, dont chaque partie et auto-décrite.

Structure de base : les paires (clé, valeur). Une **valeur** est soit une chaîne de caractère, un numérique, un Booléen, un objet ou un tableau.
Exemple codé en JSON :

```
"nom": "philippe"
```

Et le même, codé en XML :

```
<nom> philippe</nom>
```

Représentation de documents textuels

La construction (clé, valeur) permet de créer des agrégats. Voici un exemple en JSON :

```
{"nom": "Philippe Rigaux", "tél": 2157786, "email":
"philippe@gmail.com"}
```

En XML :

```
<personne>
<nom> Philippe Rigaux </nom>
<tel> 2157786 </tel>
<email> philippe@gmail.com </email>
</personne>
```