

M41L02C Cloud Computing : Le NoSQL et le BigData

Réalisé par : Julien GIRAUD, Jonathan GUICHARD, Luna NADJARE, Olivier PAUTRAT



Sommaire

1. Qu'est-ce que le NoSQL ?
 - a. SGBDs
 - b. ACID
2. NoSQL vs SQL ?
 - a. Différence de syntaxe
3. Aujourd'hui qui l'utilise ?
 - a. Le cas Discord



Qu'est-ce que le NoSQL

Plusieurs définitions :

- Utilisation d'un autre langage de requête, qui pourrait remplacer SQL
- « Not Only SQL » combiner de SQL avec d'autre mécanismes de recherche d'information

Plus de liberté au niveau de la structure de la BD : pas de normalisation

- Tableaux / listes dans les tables

Exemple : les données en JSON

```
1  {
2    "clients": [
3      {
4        "id": "59761c23b30d971669fb42ff",
5        "name": "Dunlap Hubbard"
6      },
7      {
8        "id": "59761c233d8d0f92a6b0570d",
9        "name": "Kirsten Sellers"
10     }
11   ]
12 }
```



Les SGBDs

SQL

Oracle Database

MySQL / MariaDb

SQL Server (Microsoft)

PostgreSQL

NoSQL

Apache

Cassandra

(Discord)

MongoDB

Elasticsearch

CouchDB (JSON & Api REST)



ACID

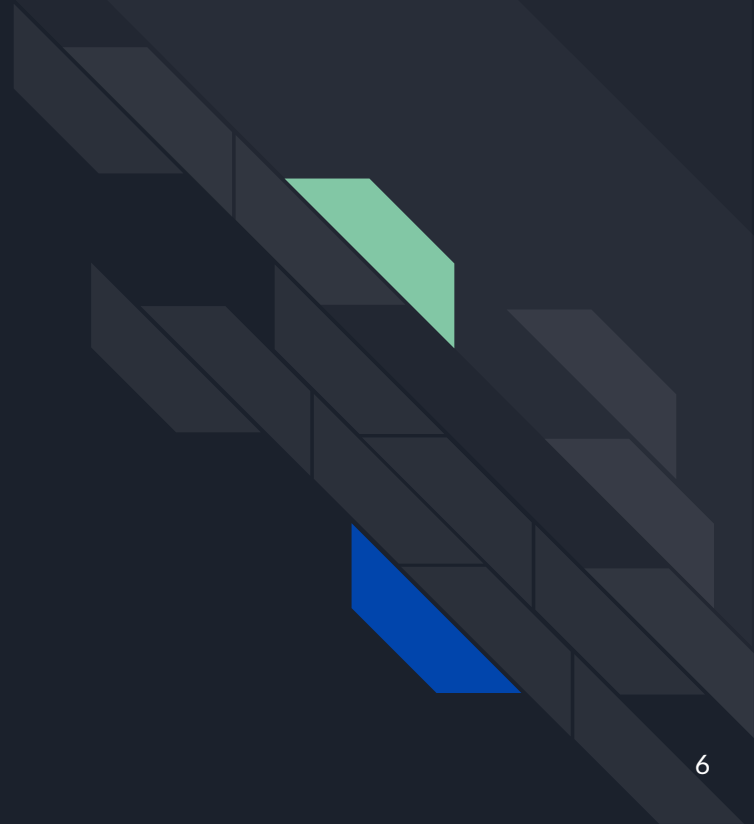
Atomicité : Une transaction se fait au complet ou pas du tout

Cohérence : Chaque transaction amènera le système d'un état valide à un autre état valide

Isolation : Toute transaction doit s'exécuter comme si elle était la seule sur le système.

Durabilité : Lorsqu'une transaction a été confirmée, elle demeure enregistrée

NoSQL vs SQL





Une différence de syntaxe

SQL

NoSQL

Select	<code>select * from emp</code>	<code>db.emp.find()</code>
Insert	<code>insert into emp (1, 'baptiste Reynaud')</code>	<code>db.emp.insert({empno : 1, name : 'baptiste R'})</code>
Update	<code>update emp set 'Baptiste Reynaud' where id = 1</code>	<code>db.emp.update({ "empno" : 1 }, { \$set: { "name" : 3 } })</code>
Create Table	<code>create table emp (empno number, name varchar)</code>	<code>db.createCollection({...})</code>

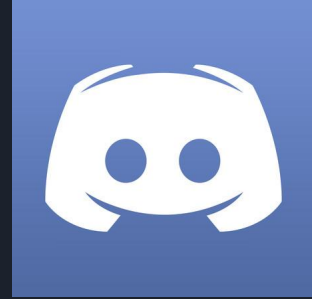


Avantages et Inconvénients

- + Les données sont distribuées sur plusieurs machines
 - + Facile d'augmenter la puissance du système
 - + Pas de goulot d'étranglement
- + Facilité à représenter des données
- Il n'existe pas de langage d'interrogation standardisé
- L'écriture de requêtes complexes est difficile à mettre en œuvre
- L'offre NoSQL est segmentée en plusieurs familles où chacune répond à un besoin précis.

Aujourd'hui qui l'utilise ?

Le cas Discord



Problèmes :

- Énormément de messages échangés (1 000 000 par ans)
- Ne pas avoir à changer de SGBD tous les ans (*scalabilité*)
- Système de gestion des erreurs (aucun crash définitif)
- Faible temps de réponse

Solution :

Cassandra (licence Apache)





Sources

<http://www.sourceamax.com/sql-vs-nosql>

<https://blog.discordapp.com/how-discord-stores-billions-of-messages>

<https://docs.mongodb.com/manual/reference/>

[https://fr.wikipedia.org/wiki/Propriétés ACID](https://fr.wikipedia.org/wiki/Propriétés_ACID)

<https://fr.wikipedia.org/wiki/Elasticsearch>