

Bases de données - NoSQL

CM 3

Les schémas de données NoSQL

Namrata PATEL

Namrata.patel@univ-montp3.fr

L'approche non relationnelle

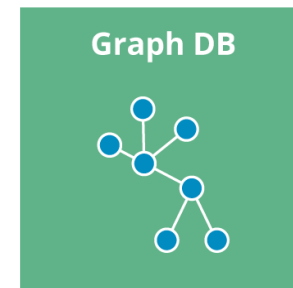
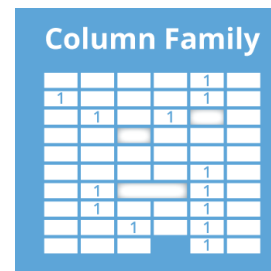
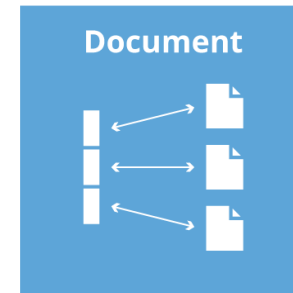
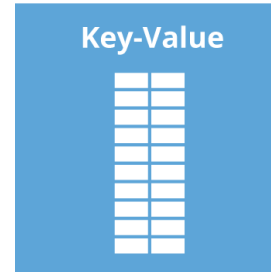
Principes, réalisation

UNE APPROCHE NON RELATIONNELLE

- « linking » ou « embedding »
 - Liaison ou duplication
- Liaison
 - Implique de gérer le déréférencement
 - Moins de redondance
- Duplication
 - Participe à l'augmentation du volume
 - Permet de gérer l'historique
- Les problèmes de jointures apparaissent avec le BigData, pas avant.

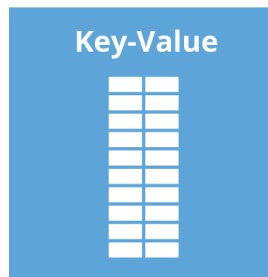
4 schémas possibles

- Orienté clé-valeurs
- Orienté documents
- Orienté colonnes
- Orienté graphes



Les 4 systèmes NoSQL

(1) Orienté Clé-Valeurs



PAIRE CLEF-VALEUR

- L'idée véhiculée par la « primary Key »
 - une clef identifie une ligne
- À partir d'une donnée simple (la clef), je veux obtenir une donnée complexe (les valeurs)

« C'est quoi son ID ? »

OÙ SONT LES CLEFS ? OÙ SONT LES VALEURS?

- Sur un réseau social, je peux obtenir
 - la liste des « amis, contacts »
 - à partir d'un **utilisateur**
- Dans un catalogue de livres, je peux obtenir
 - les détails d'un ouvrage
 - à partir de son **ISBN**
- Dans un fichier de log,
 - la **date** permet de retrouver
 - l'état du système à un instant donné

ENTREPÔT CLEF-VALEUR

- Simplicité
 - Chaque ligne peut être complètement différente
- Possibilité d'indexation secondaire
 - Seul l'accès par la clef est possible
- Ligne 1
 - Id: 101
 - Nom : X, Prénom : Y, Pays : France
- Ligne 2
 - Id: 102
 - Nom : V , Email: v@gmail.com

CONTRAINTES & REDONDANCE

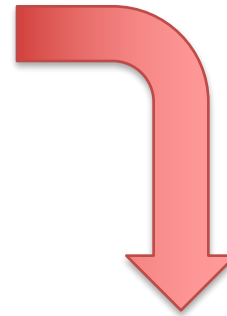
- Schema-less
- Pas de validation par le moteur NoSQL
 - La responsabilité au développeur et non au DBA
- Couteuses en espace disque

RECHERCHE SUR CRITÈRES SECONDAIRES

- Soit effectuer un scan des données
 - Potentiellement long
- Soit utiliser un index
 - Comment indexer une base schema-less ?
 - Ligne 1 Id: 101 Nom : X, Prénom : Y, Pays : France
 - Ligne 2 Id: 102 Nom : V , Email: v@gmail.com
- Index-secondaire
 - référencer en mémoire d'autres attributs que la clef

INDEXATION SECONDAIRE

Clef	Valeur
001	'Nom' : 'Toto'
002	'Nom' : 'B'
003	'Nom' : 'C'
004	'Nom' : 'Toto'



Clef	Valeur
B	'Clefs' : '002'
C	'Clefs' : '003'
Toto	'Clefs' : ['001', '004']

En résumé

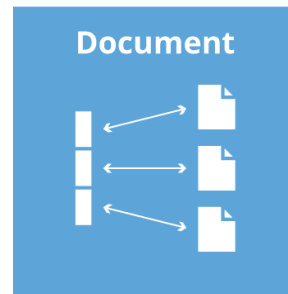
- Rapidité :
 - Stockage simple et direct d'un tableau associatif (hashtable).
- Système le plus simple :
 - Lecture, écriture et suppression.
 - Ne fait que stocker des données en rapport avec une clé.
- Pas de relation entre les données.
 - Le système ne connaît pas les données ou ce qu'elles représentent (sémantique associée).
- Interrogation par la clé.
 - On récupère une donnée associée à une clé.
- Exemples de systèmes :
 - **Riak, Project Voldemort, DynamoDB, Redis, Berkeley DB**

Exercice

- **Mise en situation avec un système clé-valeurs**
 - Contexte
 - Besoin
 - Modélisation intuitive
 - Index secondaire
 - Thm CAP : quel est votre compromis ?

Les 4 systèmes NoSQL

(2) Orienté Documents



BASES ORIENTÉES DOCUMENTS

- A la différence des clef-valeurs,
 - les valeurs stockées sont
 - des documents aux structures complexes (par exemple JSON)
- Documents **lisibles** par le moteur NoSQL

BASES ORIENTÉES DOCUMENTS

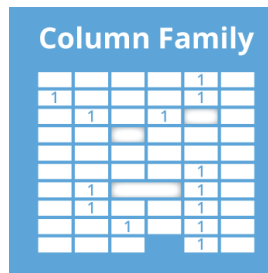
- Basé sur un système clé-valeur
- Mais la valeur est une structure que le système connaît et peut donc parcourir.
- Certains systèmes proposent les opérations croisées entres plusieurs documents.
- Exemples de systèmes :
 - **MongoDB, CouchDB, RavenDB**

Exercice

- **Mise en situation avec un système clé-valeurs**
 - Contexte
 - Besoin
 - Modélisation intuitive
 - Utilité de la valeur sous forme de document lisible
 - Thm CAP : quel est votre compromis ?

Les 4 systèmes NoSQL

(3) Orienté Colonnes



ORIENTÉE LIGNE / ORIENTÉE COLONNE

ID	NOM	PRENOM
1	Wayne	Bruce
2	Kent	Clark

Orienté « Lignes »

1	NOM	Wayne
1	PRENOM	Bruce
2	NOM	Kent
2	PRENOM	Clark

Orienté « Colonnes »

Exemple

ID	NOM	PRENOM	TEL	FAX	MAIL
1	Wayne	Bruce			batman@me.com
2	Kent	Clark			
3		Trinitiy			trinity@matrix.com

101	NOM	Wayne	PRENOM	Bruce	MAIL	batman@me.com
102	NOM	Kent	PRENOM	Clark		
103	PRENOM	Trinity	MAIL	trinity@matrix.com		

Avantages

- L'ajout de données simple et rapide :
 - les données sont concaténées les unes à la suite des autres.
- "scalable" :
 - comme le développement des données ne se fait que sur une seule dimension leurs partitionnements est plus simple à réaliser et on peut les distribuer sur plusieurs serveurs.
- Ajouter une colonne devient trivial :
 - il s'agit au final seulement d'ajouter un nouveau tuple
 - Ajouter qu'une seule colonne à une ligne en particulier est possible

En résumé

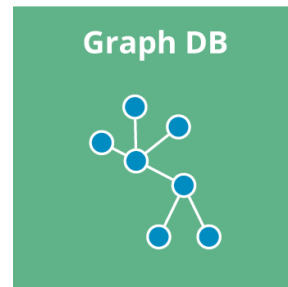
- « Opposé » d'une BD relationnelle :
 - en relationnel, le modèle est orienté n-uplets dans des tables, le nombre de colonnes est fixé.
 - Avec un modèle orienté colonnes, on ne stocke que les colonnes qui ont des valeurs non nulles regroupées dans des familles de colonnes.
- Les colonnes sont stockées sous une forme clé-valeur.
 - On peut ajouter une colonne dans un enregistrement aussi facilement que l'ajout d'un n-uplet en relationnel
- Coût de stockage d'une valeur nulle
- Stockage de plusieurs millions de colonnes
- Exemples de systèmes :
 - **BigTable, Cassandra, HBase, Hypertable**

Exercice

- **Mise en situation avec un système clé-valeurs**
 - Contexte
 - Besoin
 - Modélisation intuitive
 - Utilité de colonnes comme tuples variés
 - Thm CAP : quel est votre compromis ?

Les 4 systèmes NoSQL

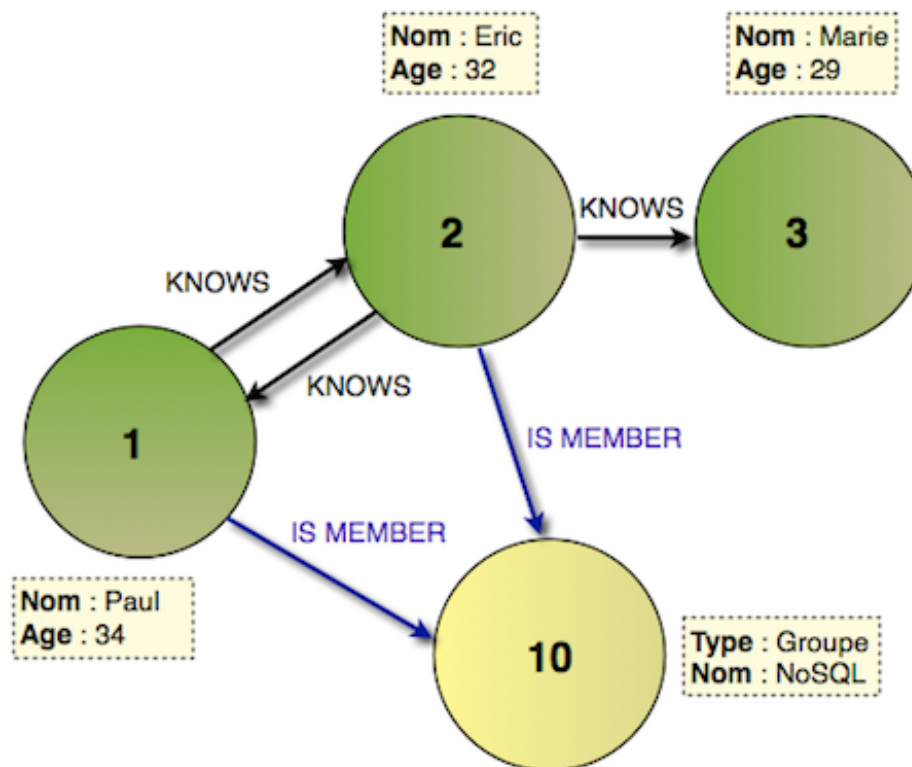
(4) Orienté graphes



BASES ORIENTÉES GRAPHES

- Base de données spécifiquement dédiée au stockage de structures de données de type graphe
 - exclusivement les données dans des noeuds et des arcs
- Cas d'école : stockage d'un réseau social.
 - Permet de parcourir les relations entre utilisateurs

Exemple



En résumé

- On ne stocke plus un simple ensemble de données mais des relations.
- Stockage d'entités et de relations entre les entités.
 - On peut ajouter des propriétés aux relations et aux entités.
- Cohérence forte des données.
 - Pas d'arête pendante.
- Difficile de monter en charge (ajout de serveurs).
- Exemples de systèmes :
 - **Neo4J, Infinite Graph, FlockDB(Twitter)**

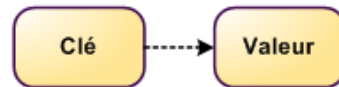
Exercice

- **Mise en situation avec un système clé-valeurs**
 - Contexte
 - Besoin
 - Modélisation intuitive
 - Utilité de liens sous forme d'arcs de graphes
 - Thm CAP : quel est votre compromis ?

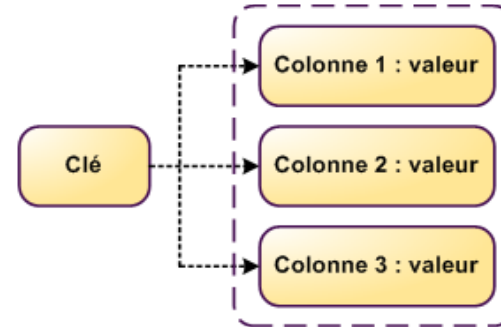
Conclusion

Résumé, comparatif

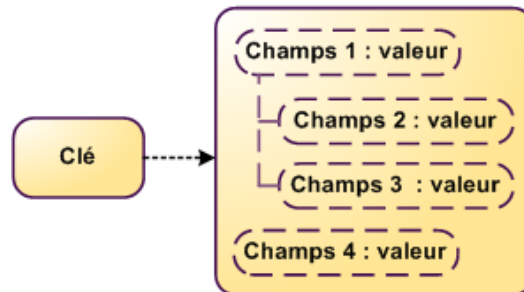
Résumé



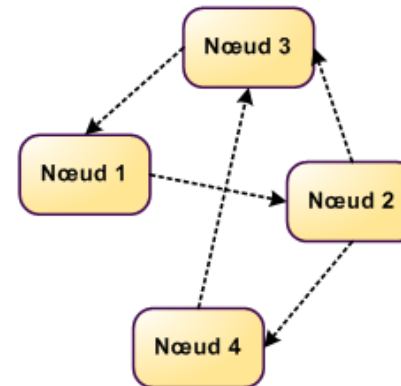
BDD Clé-Valeur



BDD Orientée colonnes



BDD Orientée document



BDD Orientée graphe

Comparatif

