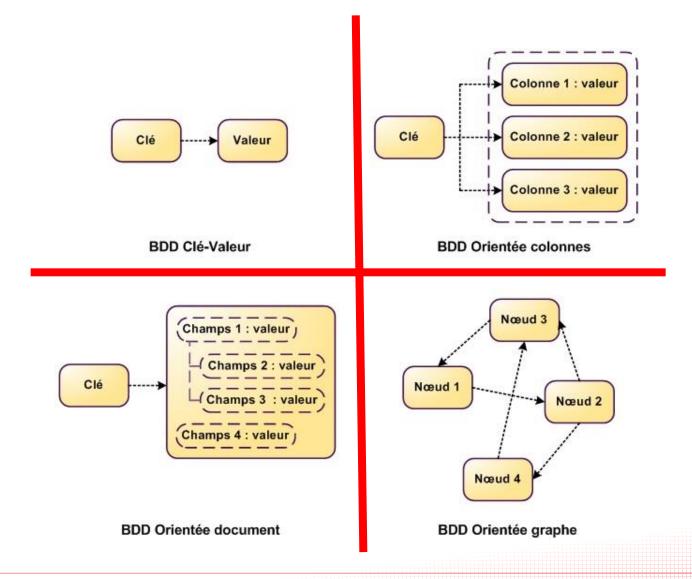
NoSQL & BigData

Ch. Stettler - HEG-Genève



Types de bdd NoSQL

Haute école de gestion Genève



Types

- 4 types différents de bdd NoSQL :
 - Orientées agrégats : Un agrégats est un string qui contient toutes sortes d'information
 - Clé-valeur
 - Documents
 - Colonnes
 - Orientées graphes :
 - Graphes

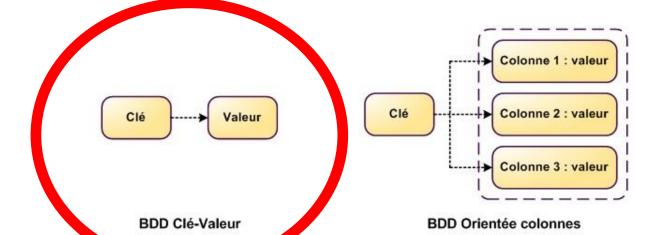
Agrégat (de données)

- Ensemble de données qui sont traitées simultanément.
 - unité d'information complexe traitée, stockée et échangée de façon atomique.
- Exemple : gestion des commandes d'un client :
 - en relationnel, on aurait une table client et une table commande (on peut donc travailler sur tous les clients, ou sur toutes les commandes).
 - en NoSQL : <u>un agrégat</u> contient UN seul client avec TOUTES SES commandes (permettant de regrouper les données à traiter ensemble).
 - ou alors UN produit avec la liste de TOUS LES clients qui l'ont commandé.
- Difficulté : définir les bons agrégats (dépend des applications).

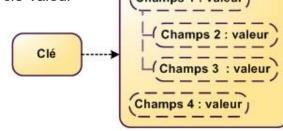
Types de bdd NoSQL

Haute école de gestion Genève

Dans la valeur, on va stocker tout un agrégat



Différence entre BDD clé-valeur et BDD docuemnt : Les champs sont toujours structurés en clé-valeur

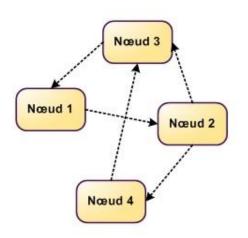


BDD clé-valeur (Redis) est le plus performant.

Donc si on veut vraiement de la performence on va utiliser Redis Mais par contre on ne peut y accéder que par la clé.

Avec MongoDB on peut chercher par les champs.

Redis stocké dans la mémoir. MongoDB stocké sur le disque. **BDD Orientée document**



BDD Orientée graphe

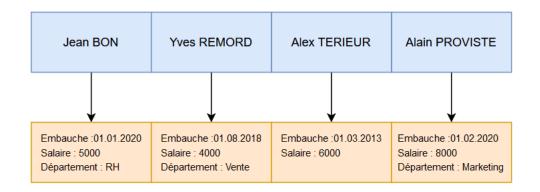
Clé-valeur

• Une clé primaire permet d'accéder à une valeur (BLOB: pas de schéma) :

Clés:

- nombre, date, texte
- photo, vidéo
- structure objet
- chaque valeur peut avoir sa structure

Valeurs:



On peut comparer cela à un dictionnaire ou un annuaire («HashMap»), si vous ne connaissez pas la clé (nom), vous ne trouverez pas la définition, ni le numéro de téléphone.

Clé-valeur : avantages

- Simplicité : valeurs peuvent être de n'importe quel type
- Rapidité : accès direct à l'information
- Évolutivité : valeurs non structurées donc facilement modifiables
- Portabilité : BDD peut être migrée facilement

Clé-valeur : inconvénients

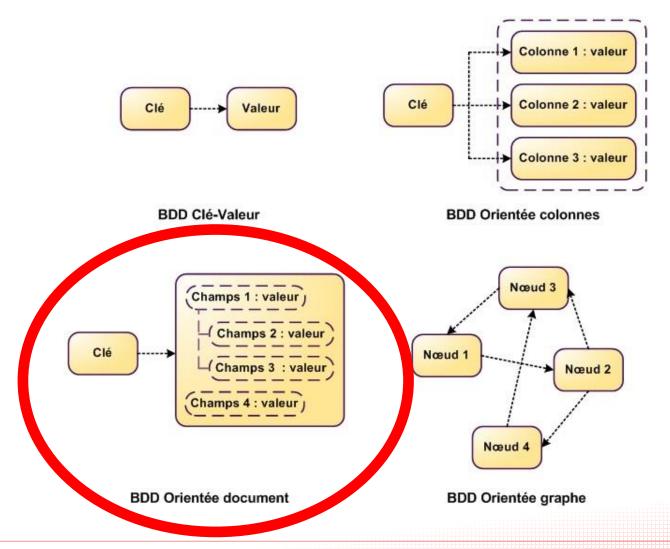
- Simplicité relative : valeurs peuvent être de n'importe quel type et deviennent compliquées à gérer
- Pas de langage de requêtes standard : le programme fait tout
- La valeur est opaque au dbms (il la considère comme un blob)
- Pas de filtres : valeur vue comme un tout.

Clé-valeur

- Exemples d'utilisation :
 - e-commerce : panier d'achat
 - chat
 - profil d'utilisateur, préférences

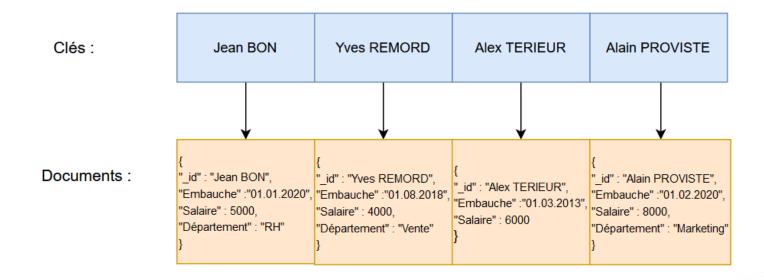
BDD:

- Redis: Trip Advisor, Nokia, Samsung, Docker
- Memcached : Wikipedia, Filckr, Wordpress
- SimpleDB : Amazon



Documents

 Repose sur le principe clé/valeur, avec une extension sur les champs qui composent ce document : la structure de l'agrégat est visible du dbms (les requêtes peuvent référencer des propriétés de l'agrégat).



Documents : avantages / inconvénients

- Avantages :
 - Modèle de données simple et puissant, structuré (XML, JSON)
 - Requêtes plus complètes
 - Flexibilité
 - Évolutif
- Inconvénients :
 - Duplication des données

Ne pas avoir peur des duplications de données. Peut-être qu'on veut accéder à une info de plusieurs façon

- Cohérence difficile
- Modèle limité à des clés

Documents

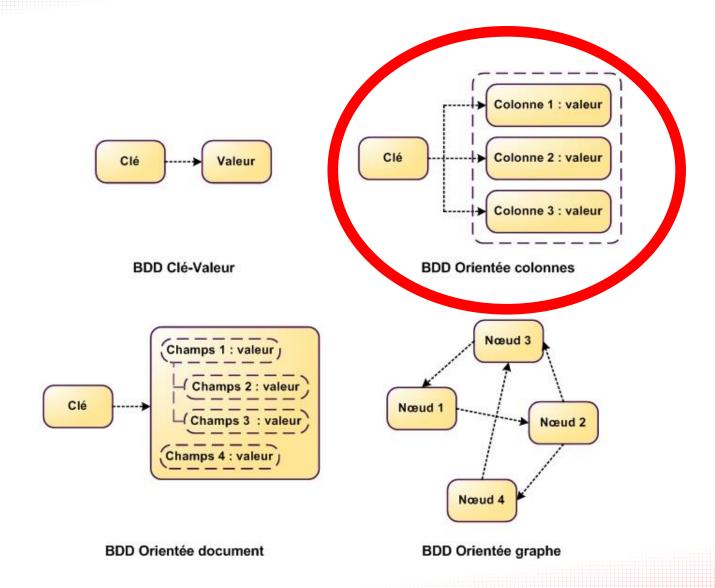
- Exemples d'utilisation :
 - Échanges interbancaires
 - Dépôts de logiciels
 - Gestion de catalogues de produits

BDD:

- MongoDB: eBay, Expedia
- CouchBase: PayPal, Ryanair
- DynamoDB : BMW

Types de bdd NoSQL

Haute école de gestion Genève



Colonnes

 Permet d'analyser un très grand volume de données du même type : récupération très simple (et rapide) de tous les noms, ou de tous les salaires, car toutes ces données (toute la colonne) sont du même type.

relationnel: orienté ligne

Id	Embauche	Salaire	Département
Jean BON	01.01.2020	5000	RH
Yves REMORD	01.08.2018	4000	Vente
Alex TERIEUR	01.03.2013	6000	
Alain PROVISTE	01.02.2020	8000	Marketing

Stockage orienté colonnes

Id	Embauche
Jean BON	01.01.2020
Yves REMORD	01.08.2018
Alex TERIEUR	01.03.2013
Alain PROVISTE	01.02.2020

Id	Salaire
Jean BON	5000
Yves REMORD	4000
Alex TERIEUR	6000
Alain PROVISTE	8000

Id	Département
Jean BON	RH
Yves REMORD	Vente
Alain PROVISTE	Marketing

Colonnes : avantages / inconvénients

- Avantages :
 - Capacité de stockage accrue
 - Accès rapide aux données
 - Flexibilité
 - Historisation à la valeur

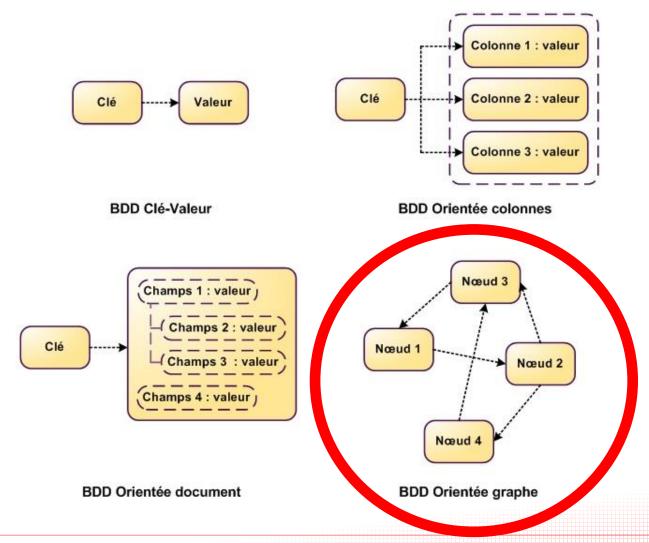
- Inconvénients :
 - Requêtes limitées

Colonnes

- Exemples d'utilisation :
 - Comptage (vote en ligne, ...)
 - Logging d'événements, suivi de colis (changement régulier du statut)
 - Statistiques

BDD:

- BigTable : Google
- Hbase: Apache, Hadoop
- Cassandra
- elasticsearch : Elastic (moteur de recherche)

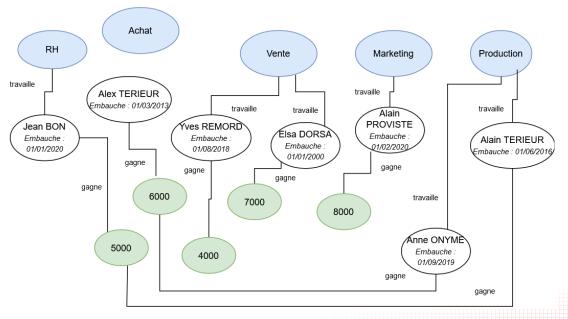


Graphes

- Gestion de nœuds (qui ont chacun leur propre structure), relations entre les nœuds, et propriétés (des nœuds et des relations)
- Facilite la représentation du monde réel

Les noeuds aussi peuvent avoir leur propre propriétés Ex. information de temps, de km

=> afin de trouver le chemin le plus court d'un noeud à un autre



Graphes: avantages / inconvénients

- Avantages :
 - Adapté aux objets complexes organisés en réseaux, avec bcp de liens
 - Permet d'appliquer les algorithmes de théorie des graphes
 - Architecture modelable
 - Très rapide pour manipuler les données fortement connectées.

- Inconvénients :
 - Architecture limitée aux données fortement connectées.

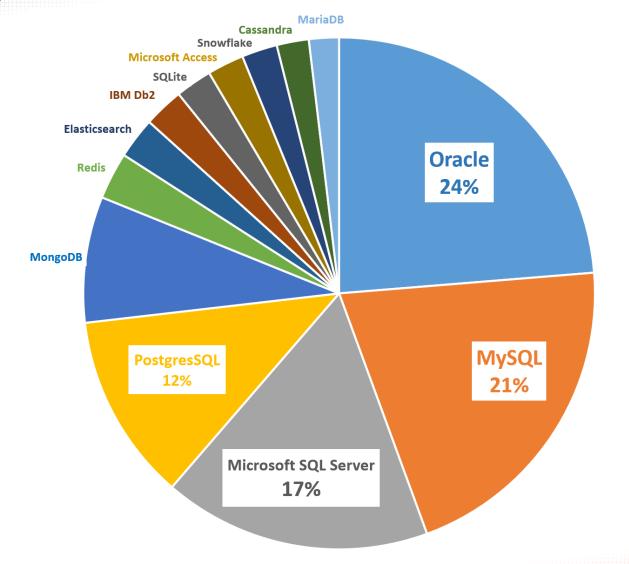
Graphes: avantages / inconvénients

- Exemples d'utilisation
 - Gestion de données relationnelles, réseaux sociaux
 - Moteurs de recommandations, intérêts communs
 - Données géospatiales, calculs d'itinéraires, plus court chemin
 - Web sémantique =>Pour l'inteligence artificielle

- BDD
 - · Neo4j

h e g

Haute école de gestion Genève



Popularité des SGBD

1	Oracle	Relational
2	MySQL	Relational
3	Microsoft SQL Server	Relational
4	PostgreSQL	Relational
5	MongoDB	Document
6	Redis	Key-value
7	Elasticsearch	Search engine
8	IBM Db2	Relational
9	SQLite	Relational
10	Microsoft Access	Relational
11	Snowflake	Relational
12	Cassandra	Wide column
13	MariaDB	Relational
14	Splunk	Search engine
15	Microsoft Azure SQL	Relational
16	Amazon DynamoDB	Multi-model info
17	Databricks	Multi-model info
18	Hive	Relational
19	Google BigQuery	Relational
20	Teradata	Relational
21	FileMaker	Relational
22	Neo4j	Graph
23	SAP HANA	Relational
24	Solr	Search engine
25	SAP Adaptive Server	Relational

Source: db-engines.com - 2023