

MASTER INFORMATIQUE DEPARTEMENT D'INFORMATIQUE (FSEA)

Année académique 2023-2024

Dr BATOUMA NARKOY Master Informatique : BDA

1

Chapitre 3

Modèles de Bases de données

Dr BATOUMA NARKOY Master Informatique : BDA

2

Introduction (1)

- Lorsqu'on est amené à construire une application web on utilise en général le modèle relationnel
- cette question ne se pose pas toutefois si on travaille avec une base de données de type **NoSQL**, comme MongoDB, Couchbase
- Nous utilisons tous finalement le **modèle relationnel** sans forcément en avoir conscience.
- Sa manière de fonctionner étant relativement naturelle.

Dr BATOUMA NARKOY Master Informatique : BDA

3

Introduction (2)

- D'ailleurs si nous revenons une étape avant, on peut légitimement s'interroger sur ce qu'est un **modèle** de base de données...
- Il s'agit simplement de **règles de construction et d'utilisation d'une base de données**.
- Un modèle est en général très **normé** afin que l'architecture d'une base reste compréhensible même en cas de changement d'équipe.
- Les bases de données avec des modèles relationnels sont encore largement utilisées

Dr BATOUMA NARKOY Master Informatique : BDA

4

Modèles de bases de données

Il existe de nombreux types de modèles de bases de données. Parmi les plus courants (1):

- **Modèle de base de données hiérarchique**
- **Modèle relationnel**
- **Modèle réseau**
- **Modèle de base de données orientée objet :**

Dr BATOUMA NARKOY Master Informatique : BDA

5

Modèles de bases de données

Il existe de nombreux types de modèles de bases de données. Parmi les plus courants (2) :

- **Modèle entité-association**
- **Modèle document**
- **Modèle entité-attribut-valeur**
- **Schéma en étoile**
- **Le modèle relationnel-objet, qui associe les deux éléments qui composent son nom**

Dr BATOUMA NARKOY Master Informatique : BDA

6

Choix d'un Modèle de bases de données

- Le choix d'un modèle dépend du système de gestion de BD que vous disposez;
- Certains SGBD sont conçus pour un seul modèle. D'autres pour plusieurs;
- Les modèles de bases de données conceptuelles généralistes sont les plus indiqués pour cartographier les relations entre les données de façon à faciliter la compréhension de ces données.
- Les modèles logiques basés sur les enregistrements sont eux davantage axés sur la manière dont les données sont stockées sur le serveur.
- Choisir un modèle en fonction de vos priorités pour la base de données: **rapidité, réduction des coûts, simplicité d'utilisation etc.**

Dr BATOUMA NARKOY Master Informatique : BDA

7

Choix d'un Modèle de bases de données

- Vous pouvez choisir de représenter une base de données selon l'un de ces modèles en fonction de plusieurs facteurs.
- Le plus important est de savoir si le système de gestion de base de données que vous utilisez prend en charge un modèle particulier.
- En effet, la plupart sont pensés pour un modèle de base de données particulier et exigent des utilisateurs qu'ils adoptent ce modèle, mais certains prennent en charge plusieurs modèles.
- En outre, tous les modèles ne s'appliquent pas à toutes les étapes du processus de conception d'une base de données.

Dr BATOUMA NARKOY Master Informatique : BDA

8

Choix d'un Modèle de bases de données

- Les modèles de bases de données conceptuelles généralistes sont les plus indiqués pour cartographier les relations entre les données de façon à faciliter la compréhension de ces données. Les modèles logiques basés sur les enregistrements sont eux davantage axés sur la manière dont les données sont stockées sur le serveur.
- Vous devez également choisir un modèle dont les points forts correspondent à vos priorités pour la base de données, **que ces priorités soient la rapidité, la réduction des coûts, la simplicité d'utilisation ou autre chose.**
- Examinons de plus près certains des modèles de bases de données les plus courants.

Dr BATOUMA NARKOY Master Informatique : BDA

9

Modèle Relationnel

- C'est le modèle le plus courant, appelé **modèle relationnel**, il utilise les **tables** pour **stocker et trier les données**;
- Les tables sont appelées des **relations**, dont chacune se compose de colonnes et de lignes.
- Chaque **colonne contient un attribut** de l'entité en question, comme le prix, le code postal ou la date de naissance.
- L'ensemble des attributs d'une relation est **appelé domaine**.

Dr BATOUMA NARKOY Master Informatique : BDA

10

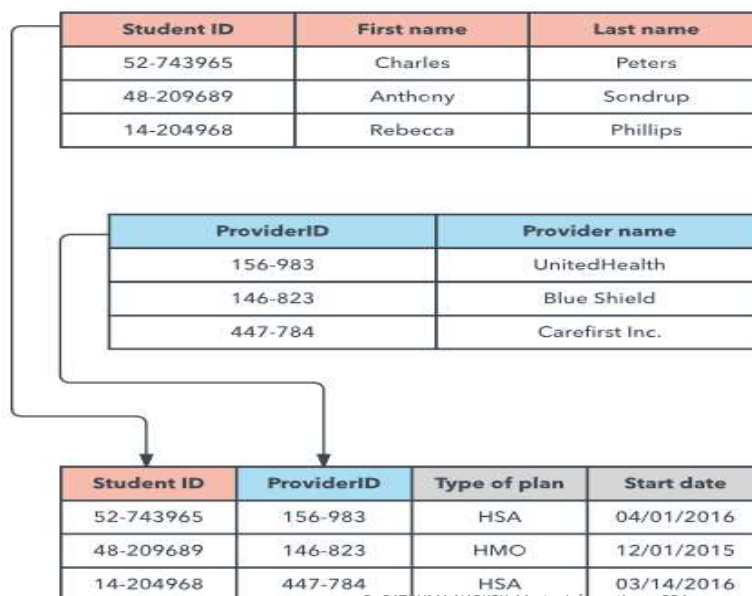
Modèle Relationnel

- La **clé primaire** est constituée par un **attribut spécifique** ou une **combinaison d'attributs**.
- On peut y faire **référence dans d'autres tables** : elle est alors appelée **clé étrangère**.
- Chaque **ligne**, également appelée **tuple**, comprend des données sur une instance spécifique de l'entité en question, comme un employé en particulier

Dr BATOUMA NARKOY Master Informatique : BDA

11

Modèle Relationnel: exemple



Dr BATOUMA NARKOY Master Informatique : BDA

12

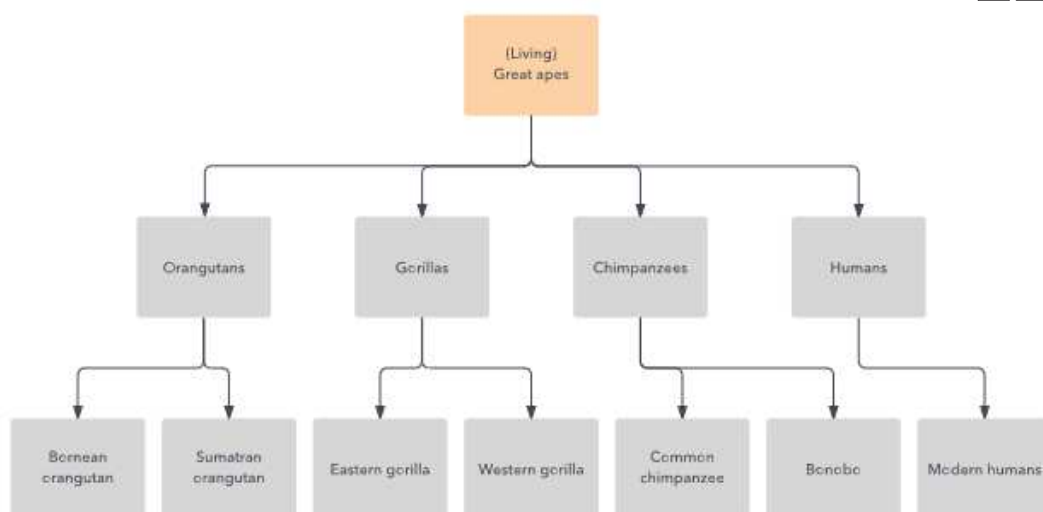
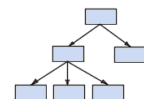
Modèle Hiérarchique

- Le **modèle hiérarchique** organise les données dans une **structure arborescente**, où chaque enregistrement dispose d'un seul parent (racine).
- Les enregistrements frères et sœurs sont triés dans un ordre particulier.
- Cet ordre est suivi pour le stockage physique de la base de données.
- Ce modèle convient à la description de nombreuses relations du monde réel.

Dr BATOUMA NARKOY Master Informatique : BDA

13

Modèle Hiérarchique



Dr BATOUMA NARKOY Master Informatique : BDA

14

Modèle Hiérarchique



- **le modèle hiérarchique** : les données sont classées hiérarchiquement, selon une arborescence descendante.
- Ce modèle utilise des pointeurs entre les différents enregistrements.
- Il s'agit du premier modèle de SGBD
- Une base de données hiérarchique est une forme de système de gestion de base de données qui lie des enregistrements dans une structure arborescente de façon à ce que chaque enregistrement n'ait qu'un seul possesseur (par exemple, une paire de chaussures n'appartient qu'à une seule personne).

Dr BATOUMA NARKOY Master Informatique : BDA

15

Modèle Hiérarchique

- Les structures de données hiérarchiques ont été largement utilisées dans les premiers systèmes de gestion de bases de données conçus pour la gestion des données du programme Apollo de la NASA.
- Cependant, à cause de leurs limitations internes, elles ne peuvent pas souvent être utilisées pour décrire des structures existantes dans le monde réel.

Dr BATOUMA NARKOY Master Informatique : BDA

16

Modèle Hiérarchique

- Les liens hiérarchiques entre les différents types de données peuvent rendre **très simple la réponse à certaines questions, mais très difficile la réponse à d'autres formes de questions.**
- **Si le principe de relation « 1 vers N » n'est pas respecté** (par exemple, un malade peut avoir plusieurs médecins et un médecin a, a priori, plusieurs patients), alors **la hiérarchie se transforme en un réseau.**
 -
- Il a surtout été utilisé par les systèmes de gestion d'information d'IBM dans les années 60 et 70, qui ont aujourd'hui majoritairement disparu en raison de certaines inefficacités opérationnelles.

Dr BATOUMA NARKOY Master Informatique : BDA

17

Avantages du Modèle Hiérarchique

- Adéquation du modèle avec les entreprises à structure arborescente
 - un grand nombre d'organisations économiques et sociales correspondent à ce modèle.
- Simplicité du modèle et implémentation facile.
- Adéquation de la structure du schéma et des besoins des utilisateurs.

Dr BATOUMA NARKOY Master Informatique : BDA

18

Inconvénients du Modèle Hiérarchique

- Impossibilité de représenter directement les liens **N : M** qui entraîne des redondances.
- Anomalies pour les opérations de stockage :
 - La suppression d'un nœud entraîne la disparition des descendants
 - ex. suppression du pilote n° 2 -> disparition des informations concernant la CARAVELLE n° 4 ou le VOL IT109).
 - L'insertion d'une information impose la création d'un segment parent.
 - Le remplacement d'une information doit s'effectuer pour toutes ses occurrences dans des arbres distincts.
 -
- Chemin d'accès unique aux données.
- **Indépendance logique très réduite.**

Dr BATOUMA NARKOY Master Informatique : BDA

19

Modèle Réseau

- Le modèle réseau **est une extension du modèle hiérarchique qui autorise des relations plusieurs-à-plusieurs entre des enregistrements liés**, ce qui implique plusieurs enregistrements parents.
- Basé sur la théorie mathématique des ensembles, ce modèle s'articule autour d'ensembles d'enregistrements connexes.
- Chaque ensemble se compose d'un enregistrement propriétaire (ou enregistrement parent) et d'un ou plusieurs enregistrements membres (ou enfants).
- Un enregistrement peut être un membre ou un enfant dans plusieurs ensembles, ce qui permet à ce modèle de traduire des relations complexes.

Dr BATOUMA NARKOY Master Informatique : BDA

20

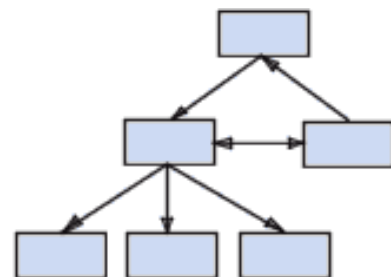
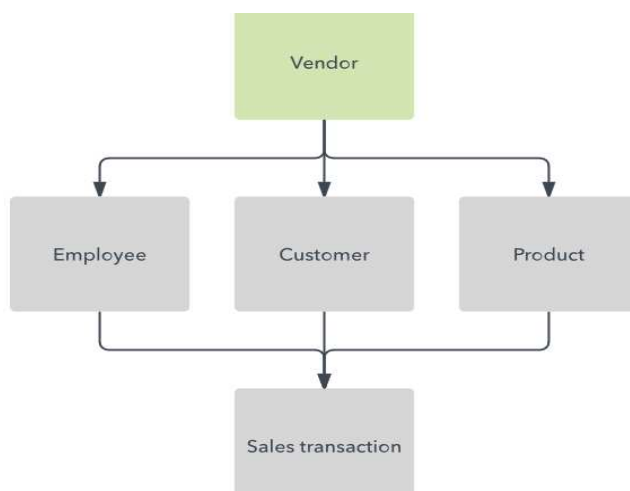
Modèle Réseau

- Son pic de popularité remonte aux années 70, après qu'il a été officiellement défini par la conférence sur les langages de systèmes de traitement de données (Conference on Data Systems Languages, CODASYL)
- Le modèle réseau est en mesure de lever de nombreuses difficultés du modèle hiérarchique grâce à la possibilité d'établir des liaisons de type n-n, les liens entre objets pouvant exister sans restriction.
- Pour retrouver une donnée dans une telle modélisation, il faut connaître le chemin d'accès (les liens) ce qui rend les programmes dépendants de la structure de données
- Ce modèle de bases de données a été inventé par C.W. Bachman. Pour son modèle, il reçut en 1973 le prix Turing.

Dr BATOUMA NARKOY Master Informatique : BDA

21

Modèle Réseau



Dr BATOUMA NARKOY Master Informatique : BDA

22

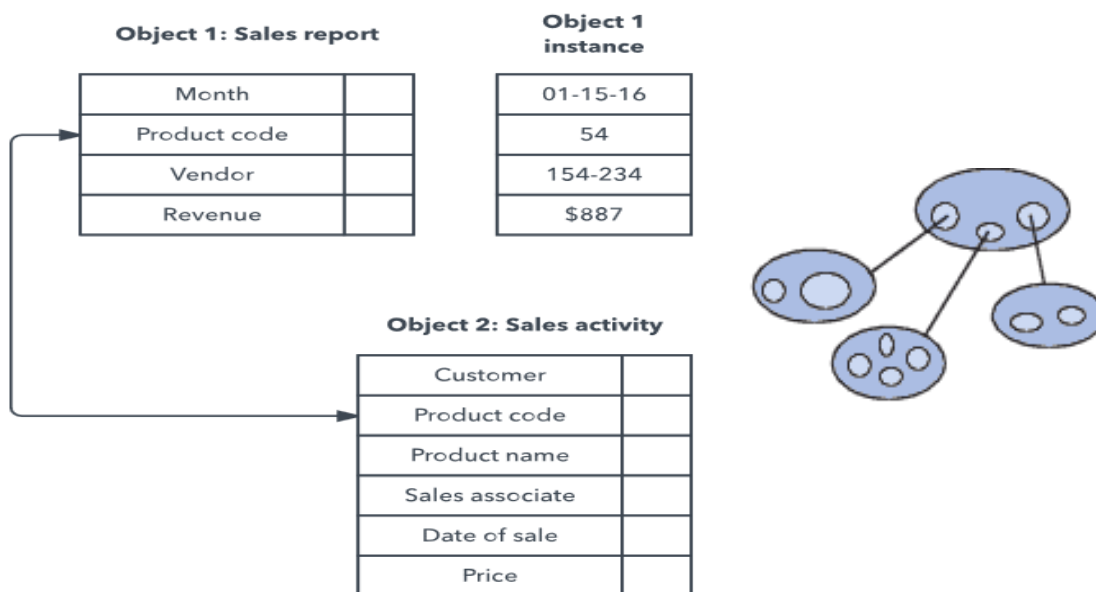
Modèle de BD Orienté Objet

- Ce modèle définit une base de données comme une **collection d'objets**, ou **d'éléments logiciels réutilisables**, associés à **des caractéristiques et des méthodes**.
- Il existe plusieurs types de bases de données orientées objet :
 - Une **base de données multimédia** réunit des médias, tels que des images, qui ne peuvent pas être stockés dans une base de données relationnelle.
 - Une **base de données hypertextuelle** permet à n'importe quel objet d'être relié à un autre. Elle est utile pour organiser de nombreuses données disparates, mais peu adaptée à l'analyse numérique.

Dr BATOUMA NARKOY Master Informatique : BDA

23

Modèle de BD Orienté Objet



Dr BATOUMA NARKOY Master Informatique : BDA

24

Modèle Relationnel-Objet

- Ce modèle de base de données **hybride** associe la **simplicité du modèle relationnel** à certaines des fonctionnalités avancées du modèle de la base de données **orientée objet**.
-
- Il permet aux concepteurs **d'intégrer des objets dans la structure bien connue des tables**.
- Les langages et interfaces d'appel comprennent **SQL3**, les langages des fournisseurs, ODBC, JDBC et les interfaces d'appel propriétaires qui sont des extensions des langages et des interfaces utilisés par le modèle relationnel

Dr BATOUMA NARKOY Master Informatique : BDA

25

Modèle Relationnel-Objet

Pourquoi étendre le modèle relationnel ?

- La reconstitution d'objets complexes éclatés sur plusieurs tables relationnelles est coûteuse car elle occasionne de nombreuses jointures pour échapper aux éclatements-jointures
- L'Objet Relationnel réhabilite les références qui permettent d'implanter des structures complexes, les attributs multivalués (tableaux, ensembles ou listes)
- Même si ce problème n'est pas inhérent au modèle relationnel, SQL92 ne permet pas de créer de nouveaux types, ce qui implique un manque de souplesse et une interface difficile avec les applications orientées objet
- L'Objet Relationnel (et SQL99) permet de définir de nouveaux types utilisateur simples ou complexes (User data type), avec des fonctions ou procédures associées comme dans les classes des langages objet
- L'Objet Relationnel supporte l'héritage de type pour profiter du polymorphisme et faciliter la réutilisation

Dr BATOUMA NARKOY Master Informatique : BDA

26

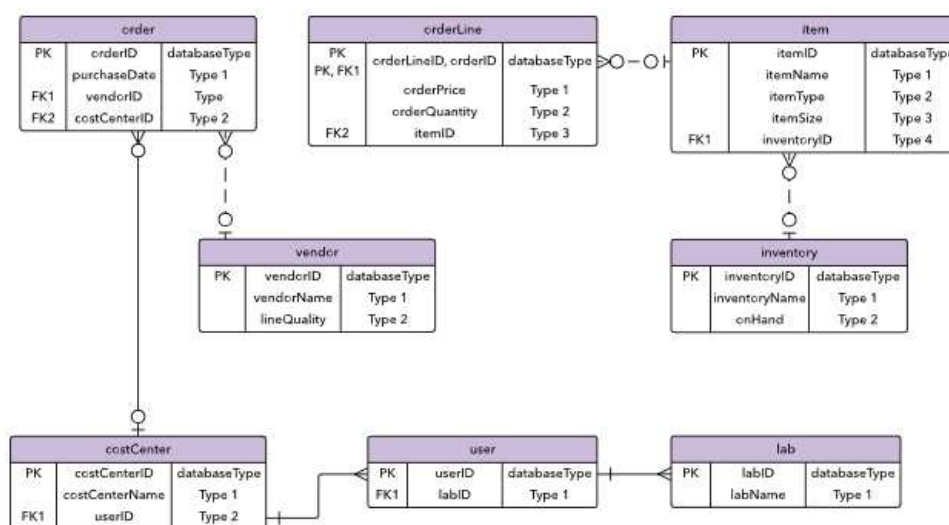
Modèle Entité-Association

- Ce modèle reproduit les **relations** entre les **entités du monde réel** de façon très similaire au modèle réseau, mais il n'est pas lié aussi directement à la structure physique de la base de données.
- Il est souvent utilisé pour la **création d'une base de données d'un point de vue conceptuel**.
- Ici, les personnes, les lieux et les objets à propos desquels les points de données sont stockés sont appelés **entités**;
- chacune d'entre elles possédant certains **attributs** qui, ensemble, composent leur **domaine**.
- On schématise aussi la **cardinalité** (relations entre les entités).

Dr BATOUMA NARKOY Master Informatique : BDA

27

Modèle Entité-Association

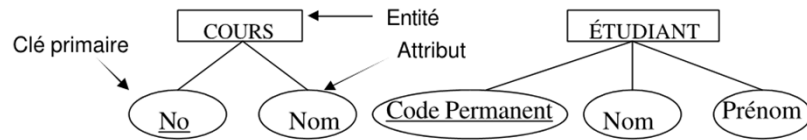


Dr BATOUMA NARKOY Master Informatique : BDA

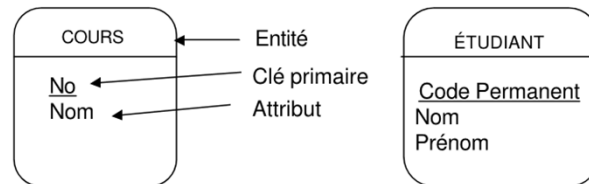
28

Représentation graphique

1ère méthode:



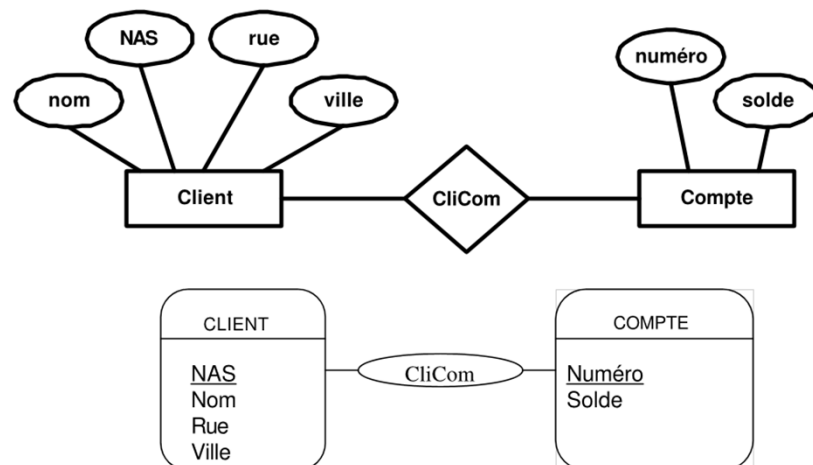
2ème méthode:



29

Représentation graphique

Ex : étudiant Claveau " Suit le cours " de Base de données
client Olivier " Possède " le compte #259



30

Modèle de Fichier inverser

- Une base de données créée selon **une structure de fichier inversé est conçue pour accélérer les recherches en texte intégral.**
- Dans ce modèle, le contenu des données **est indexé sous la forme d'une série de clés dans une table de recherche**, dont les valeurs indiquent l'emplacement des fichiers associés.
- Cette structure peut fournir des rapports quasi instantanés dans le domaine des **Big Data** ou de **l'analytique**, par exemple.
- Ce modèle est utilisé par le système de gestion de base de données ADABAS de Software AG depuis 1970, et il est encore pris en charge aujourd'hui.

Dr BATOUMA NARKOY Master Informatique : BDA

31

Modèle de Fichier inverser

- Les index inversés sont notamment utilisés pour des applications de [moteur de recherche](#).
- Par exemple, si nous effectuons une recherche des termes "ceci" et "est", alors la ou les phrases correspondantes peuvent être retrouvées par l'intersection des index des mots correspondant (le critère d'indexation est la présence du mot dans la phrase)
- Organisation inversée : principe
 - Fichier inversé = fichier index é + index secondaire(s).
 - Permet d'accéder aux enregistrements du fichier à partir de la valeur de champs autres que la clé
 - Ces champs servent de "clé" aux index secondaires (d'où le nom de Clé secondaire)

Dr BATOUMA NARKOY Master Informatique : BDA

32

Modèle de BD orienté texte

- Le modèle de base de données orientée texte est **le plus ancien et le plus simple**.
- Il énumère simplement **toutes les données dans une seule table**, composée de **colonnes et de lignes**.
- Pour **accéder aux données ou les manipuler**, l'ordinateur doit lire **l'intégralité du fichier plat dans la mémoire**, ce qui rend ce modèle **inefficace**, sauf pour les ensembles de données les plus petits.

Dr BATOUMA NARKOY Master Informatique : BDA

33

Modèle Multidimensionnel

- Avec le besoin grandissant de **stocker toujours plus de données et d'y accéder de manière toujours plus rapide et pertinente**, de nouveaux modèles ont été développés parmi lesquels le modèle multidimensionnel;
- La multi dimensionnalité suppose ici qu'un tableau, tel que connu dans le modèle relationnel, a plusieurs dimensions, là où un tableau relationnel n'en a que deux (colonne, ligne);
- Plutôt que de représenter les données sous forme de tableaux, le modèle multidimensionnel travaille avec plusieurs "axes", le plus souvent 3, formant ce qu'on appelle un hypercube;
- Il s'agit d'une variante du modèle relationnel conçue pour améliorer le traitement analytique;
- Alors que le modèle relationnel est optimisé pour le traitement transactionnel en ligne (OLTP), ce modèle est conçu pour le traitement analytique en ligne (OLAP).

Dr BATOUMA NARKOY Master Informatique : BDA

34

Modèle Multidimensionnel

- Chaque cellule d'une base de données dimensionnelle contient des données sur les dimensions suivies par la base de données.
- Visuellement, elle ressemble à un ensemble de cubes plutôt qu'à des tables en deux dimensions.
- En réalité, **une base multidimensionnelle est contenue dans une seule table**,
- chaque cellule, ou "fait", étant par ailleurs caractérisée par une dimension (contexte du fait : type, date, lieu, groupe) et une mesure (quantité descriptive).

Dr BATOUMA NARKOY Master Informatique : BDA

35

Modèle Multidimensionnel

- Les bases multidimensionnelles sont le plus souvent **formées par agrégats de bases** pouvant être relationnelles, en tout cas hétérogènes.
- Les données ainsi agrégées peuvent être analysées avec un nouveau type d'outil, OLAP (*On-Line Analytical Processing*) - là où les bases relationnelles utilisent OLTP (*On-Line Transaction Processing*).
- L'OLAP permet une analyse et une visualisation des données plus fines, pouvant utiliser **plusieurs niveaux de granularité**.
- **La complexité d'OLAP le limite à une utilisation dans le cas de requêtes de grande envergure.**
- Certaines de ses commandes SQL divergent largement de celles dont on a l'habitude : RANK(), DENSE_RANK(), NTILE(), PERCENT_RANK()...

Dr BATOUMA NARKOY Master Informatique : BDA

36

OLTP VS OLAP

- Le traitement analytique en ligne (OLAP) et le traitement des transactions en ligne (OLTP) sont des systèmes de traitement de données qui vous aident à stocker et à analyser des données métier.
- Vous pouvez collecter et stocker des données provenant de sources multiples, telles que des sites web, des applications, des compteurs intelligents et des systèmes internes.
- OLAP combine et regroupe les données afin que vous puissiez les analyser de différents points de vue.
- À l'inverse, OLTP stocke et met à jour les données transactionnelles à volume élevé de manière fiable et efficace. Les bases de données OLTP peuvent être l'une des nombreuses sources de données d'un système OLAP.

Dr BATOUMA NARKOY Master Informatique : BDA

37

Similitudes: entre OLTP et OLAP

- Le traitement analytique en ligne (OLAP) et le traitement des transactions en ligne (OLTP) sont des systèmes de gestion de bases de données permettant de stocker et de traiter des données en grandes quantités. Ils nécessitent une infrastructure informatique efficace et fiable pour fonctionner correctement.
- Vous pouvez les utiliser à la fois pour interroger des données existantes ou pour stocker de nouvelles données. Les deux aident l'organisation à prendre des décisions fondées sur des données.
- La plupart des entreprises utilisent conjointement des systèmes OLTP et OLAP pour répondre à leurs besoins en matière d'informatique décisionnelle. Cependant, l'approche et l'objectif de la gestion des données diffèrent considérablement entre OLAP et OLTP.

Dr BATOUMA NARKOY Master Informatique : BDA

38

Différences : OLTP par rapport à OLAP

- L'objectif principal du traitement analytique en ligne (OLAP) est d'analyser des données agrégées, tandis que celui du traitement des transactions en ligne (OLTP) est de traiter les transactions de base de données.
- Vous utilisez les systèmes OLAP pour générer des rapports, effectuer des analyses de données complexes et identifier des tendances. En revanche, vous utilisez des systèmes OLTP pour traiter les commandes, mettre à jour l'inventaire et gérer les comptes clients.
- Les autres différences majeures concernent la mise en forme des données, l'architecture des données, les performances et les exigences. Nous aborderons également un exemple de cas où une organisation peut utiliser OLAP ou OLTP.

Dr BATOUMA NARKOY Master Informatique : BDA

39

Exemple de comparaison entre OLTP et OLAP

- Prenons l'exemple d'une grande entreprise de vente au détail qui exploite des centaines de magasins à travers le pays. L'entreprise dispose d'une énorme base de données qui permet de suivre les ventes, les stocks, les données clients et d'autres indicateurs clés.
- L'entreprise utilise OLTP pour traiter les transactions en temps réel, mettre à jour les niveaux d'inventaire et gérer les comptes clients. Chaque magasin est connecté à la base de données centrale, qui met à jour les niveaux d'inventaire en temps réel au fur et à mesure que les produits sont vendus.
- L'entreprise utilise également OLTP pour gérer les comptes clients, par exemple pour suivre les points de fidélité, gérer les informations de paiement et traiter les retours.

Dr BATOUMA NARKOY Master Informatique : BDA

40

Exemple de comparaison entre OLTP et OLAP

- Parallèlement, la société utilise OLAP pour analyser les données collectées par OLTP.
- Les analystes commerciaux de l'entreprise peuvent utiliser OLAP pour générer des rapports sur les tendances des ventes, les niveaux d'inventaire, les données démographiques des clients et d'autres indicateurs clés.
- Ils effectuent des requêtes complexes sur de grands volumes de données historiques afin d'identifier les modèles et les tendances susceptibles d'éclairer les décisions commerciales.
- Ils identifient les produits populaires au cours d'une période donnée et utilisent ces informations pour optimiser les budgets d'inventaire.

Dr BATOUMA NARKOY Master Informatique : BDA

41

Modèle Semi-structuré

- Dans ce modèle, **les données structurelles habituellement contenues dans le schéma de base de données sont intégrées aux données elles-mêmes.**
- La **distinction** entre les **données** et le **schéma** est donc pour le moins vague.
- Ce modèle est utile pour décrire les systèmes, comme certaines sources de données sur le Web, que l'on traite comme des bases de données, **mais qui ne peuvent pas s'adapter à un schéma.**
- Il est également utile pour décrire des **interactions entre des bases de données qui ne respectent pas le même schéma.**

Dr BATOUMA NARKOY Master Informatique : BDA

42

Données structurées

- Les données structurées sont **des données dont les éléments sont adressables pour une analyse efficace**. Il constitue 20% du total des données.
- Il est organisé dans un **référentiel formaté** qui est généralement **une base de données**. Il concerne toutes les données pouvant être stockées dans une base de données SQL dans une table avec des lignes et des colonnes.
- Ils ont une clé relationnelle et peuvent facilement être mappés dans des champs prédéfinis.
- Aujourd'hui, ces données sont le plus souvent traitées lors du développement et constituent le moyen le plus simple de gérer les informations.

ID	Name	Age
1	Alex	22
2	Bob	24
3	Emily	32

Dr BATOUMA NARKOY Master Informatique : BDA

43

Données semi-structurées

- Les **données semi-structurées** sont des **informations qui ne résident pas dans une base de données rationnelle**,
- mais qui **possèdent des propriétés organisationnelles facilitant leur analyse**.
- Avec certains processus, vous pouvez les stocker dans la base de données relationnelle (cela peut être très difficile pour certaines données semi-structurées), mais Semi-structurée existe pour réduire l'espace.

Exemple: données XML

```

1. <?xml version = "1.0" encoding="UTF-8" standalone="yes"?>
2. <document>
3.   <employee>
4.     <name>Alex</name>
5.     <age>22</age>
6.   </employee>
7.   <employee>
8.     <name>Bob</name>
9.     <age>24</age>
10.  </employee>
11.  <employee>
12.    <name>Emily</name>
13.    <age>32</age>
14.  </employee>
15. </document>

```

Dr BATOUMA NARKOY Master Informatique : BDA

44

Données non structurées

- Les **données non structurées** sont des données qui **ne sont pas organisées de manière prédéfinie** ou qui **ne possèdent pas de modèle de données prédéfini**.
- Elles **ne conviennent donc pas à une base de données relationnelle traditionnelle**.
- Ainsi, pour les données non structurées, il existe des plates-formes alternatives pour le stockage et la gestion.
- Elles sont de plus en plus répandues dans les systèmes informatiques et sont utilisées par les entreprises dans diverses applications de business intelligence et d'analyse.

Dr BATOUMA NARKOY Master Informatique : BDA

45

Comparaison

Exemple: Word, PDF, texte, logs.

Table de comparaison

Structurées	Semi-structurées	Non structurées
Il est basé sur les tables de base de données relationnelle	Il est basé sur XML/RDF	Il est basé sur des caractères et des données binaires
Il est dépendant du schéma et moins flexible	Il est plus flexible que les données structurées mais moins que les données non structurées	Très flexible et l'absence de schéma
Il est très difficile de mettre à l'échelle le schéma de base de données	La mise à l'échelle est plus simple que les données structurées	C'est très facile à mettre à l'échelle

Dr BATOUMA NARKOY Master Informatique : BDA

46

Modèle Contextuel

- Ce modèle peut incorporer des éléments provenant d'autres modèles de bases de données selon les besoins.
- Il réunit des éléments des modèles orienté objet, semi-structuré et réseau.

Dr BATOUMA NARKOY Master Informatique : BDA

47

Modèle d'Association

Le modèle d'association structure les données en deux ensembles :

- Un **ensemble d'éléments**, chacun ayant un identificateur unique, un nom et un type
- Un **ensemble de liens**, chacun ayant un identificateur unique et les identificateurs uniques d'une source, d'un verbe et d'une cible.
- Le fait stocké est en lien avec la source, et chacun des trois identificateurs peut faire référence soit à un lien, soit à un élément.

Dr BATOUMA NARKOY Master Informatique : BDA

48

Le Big data

- Le **Big Data** désigne les mégadonnées collectées par les entreprises de toutes les industries, analysées afin d'en dégager de précieuses informations.
- Avant de définir le Big Data, ou les mégadonnées, il est important de **bien comprendre ce que sont les données**. Ce terme définit les quantités, les caractères ou les symboles sur lesquels des opérations sont effectuées par un ordinateur.
- Les données peuvent être stockées ou transmises sous forme de signaux électriques et enregistrées sur un support mécanique, optique ou magnétique.
- Le **terme de Big Data** désigne de vastes ensembles de données collectées par les entreprises, pouvant être explorées et analysées afin d'en dégager des informations exploitables ou utilisées pour des projets de **Machine Learning**.

Dr BATOUMA NARKOY Master Informatique : BDA

49

Le Big data

- On définit souvent le Big Data par les » 3 V » qui le caractérisent : le **volume** et la **variété des données**, et la **vélocité avec laquelle elles sont générées**, collectées et traitées. C'est ce qui différencie les » mégadonnées » des données traditionnelles.
- Ces trois caractéristiques furent identifiées **pour la première fois en 2001** par Doug Laney, analyste chez Meta Group Inc. Elles furent ensuite popularisées par Gartner suite à l'acquisition de Meta Group en 2005.
- De nos jours, on attribue parfois d'autres caractéristiques au Big Data comme la véracité, la valeur et la variabilité.
- Dans les entreprises de toutes les industries, les systèmes permettant de traiter et de stocker le Big Data sont devenus indispensables. Pour cause, **les outils traditionnels** de gestion de données ne sont pas en mesure de stocker ou de traiter de tels ensembles massifs.

Dr BATOUMA NARKOY Master Informatique : BDA

50

A quoi sert le Big data?

- Dans tous les secteurs, les entreprises utilisent le Big Data engrangé dans leurs systèmes à différentes fins.
- Il peut s'agir:
 - d'améliorer les opérations;
 - de proposer un meilleur service client;
 - de créer **des campagnes marketing personnalisées** basées sur les préférences des consommateurs;
 - ou tout simplement d'augmenter le chiffre d'affaires.
- Grâce au Big Data, les entreprises peuvent profiter d'un **avantage compétitif** face à leurs concurrents n'exploitant pas les données. Elles peuvent prendre des décisions plus rapides et plus précises, s'appuyant directement sur les informations.

Dr BATOUMA NARKOY Master Informatique : BDA

51

A quoi sert le Big data?

- Par exemple, une entreprise peut analyser le Big Data pour découvrir de précieuses informations sur les besoins et **les attentes de ses clients**.
- Ces informations peuvent ensuite être exploitées pour créer de nouveaux produits ou des campagnes marketing ciblées afin d'accroître la fidélité client ou d'augmenter le taux de conversion. Une entreprise s'appuyant totalement sur les données pour aiguiller son évolution est qualifiée de « data-driven » (dirigée par les données).
- En outre, le Big Data est utilisé dans le domaine **de la recherche médicale**. Il permet notamment d'identifier des facteurs de risque de maladies, ou de réaliser des diagnostics plus fiables et plus précis.
- Les données médicales permettent aussi d'anticiper et de suivre les éventuelles épidémies.

Dr BATOUMA NARKOY Master Informatique : BDA

52

A quoi sert le Big data?

- Les mégadonnées sont utilisées dans presque tous les secteurs sans exception.
- L'**industrie de l'énergie** s'en sert pour découvrir des zones de forage potentielles et surveiller leurs opérations ou le réseau électrique.
- Les services financiers l'utilisent pour gérer les risques et analyser les données du marché en temps réel.
- Les fabricants et **les entreprises de transport**, quant à eux, gèrent leurs chaînes logistiques et optimisent leurs itinéraires de livraison grâce aux données.
- De même, les gouvernements exploitent le Big Data pour la prévention du crime ou pour les initiatives de Smart City.

Dr BATOUMA NARKOY Master Informatique : BDA

53

Les différents types de Big data

- Les données du Big Data proviennent de sources diverses, et peuvent donc prendre plusieurs formes. On distingue **plusieurs catégories principales**.
- Lorsque les données pouvant être stockées et traitées dans un format fixe et bien défini, on parle alors de **données structurées**.
- Grâce aux nombreuses avancées réalisées dans le domaine de l'informatique, des techniques permettent aujourd'hui de travailler efficacement avec ces données et d'en dégager toute la valeur.
- Cependant, même les données structurées peuvent poser problème **à cause de leur volume massif**. Alors que le volume d'un ensemble atteint désormais plusieurs zettabytes, le stockage et le traitement représentent de véritables défis.

Dr BATOUMA NARKOY Master Informatique : BDA

54

Les différents types de Big data

- Les données dont le format ou la structure sont inconnus, quant à elles, sont considérées comme des **données non structurées** .
- Ce type de données présente de nombreux défis en termes de traitement et d'exploitation, au-delà de leur volume massif.
- En guise d'exemple typique, on peut évoquer **une source de données hétérogène** contenant une combinaison de fichiers textuels, d'images et de vidéo.
- À l'ère du numérique et du multimédia, ce type de données est de plus en plus fréquent. Les entreprises ont donc de vastes quantités de données à portée de main, mais peinent à en profiter à cause de la difficulté à traiter ces informations non structurées...

Dr BATOUMA NARKOY Master Informatique : BDA

55

Les différents types de Big data

- Enfin, les **données » semi-structurées «** sont à mi-chemin entre ces deux catégories.
- Il peut s'agir par exemple de données structurées en termes de format, mais n'étant pas clairement définies au sein d'une base de données.
- Avant de pouvoir traiter et analyser les données non structurées ou semi-structurées, il est nécessaire de les préparer et de les transformer grâce à différents types d'outils de data mining ou de préparation de données.

Dr BATOUMA NARKOY Master Informatique : BDA

56

Les techniques d'analyse du Big data

- L'**analyse comparative** permet par exemple à une entreprise de comparer les performances de ses produits et services auprès des clients avec ceux de ses concurrents.
- L'**analyse marketing** consiste à analyser les données permettant de faire la promotion de nouveaux produits et services de manière mieux informée et innovante.
- L'**analyse de sentiment** a pour but d'évaluer la satisfaction client à l'égard d'une marque, notamment en passant en revue les critiques ou les commentaires laissés sur internet.
- Dans la même optique, l'**analyse des réseaux sociaux** permet de mettre en lumière la réputation d'une entreprise à partir de ce que les internautes en disent sur les réseaux. Il devient alors possible d'identifier de nouvelles audiences cibles pour les campagnes de marketing.

Dr BATOUMA NARKOY Master Informatique : BDA

57

Les Modèles NoSQL

- La principale différence avec le modèle SQL vient dans l'**absence d'un modèle relationnelle**, ce qui signifie que les données ne sont plus matérialisées et représentées par des tables.
- Cette **absence** (ou flexibilité, en fonction des modèles) **de schéma de données** permet de faire évoluer l'architecture de la base de données avec le temps, ce qui est beaucoup plus difficile dans un contexte SQL.
- Les bases de données NoSQL sont également pensées pour la **mise à l'échelle**. Elles ont la possibilité d'être **distribuées**, ce qui permet à la fois de supporter de grandes charges de travail (milliers de lectures/écritures par seconde) tout en assurant une haute disponibilité, grâce aux mécanismes de réplication de données.

Dr BATOUMA NARKOY Master Informatique : BDA

58

Les Modèles NoSQL

Parmi les bases dites NoSQL, nous retrouvons plusieurs modèles de bases de données:

- Les bases orientées **colonnes** : Cassandra, AWS DynamoDB, HBase.
 - Les bases orientées **documents** : MongoDB, Elasticsearch.
 - Les bases orientées **clé/valeur** : Redis, Memcached.
 - Les bases orientées **graphes** : Neo4j, InfluxDB.
- Le choix du modèle de base NoSQL dépend principalement des besoins et du cas d'usage. Si l'on doit souvent manipuler des données temporelles, on se dirigera vers une base adaptée pour les **séries temporelles**. À l'inverse, les bases **orientées colonnes** peuvent être intéressants dans les situations où l'on doit gérer des articles ECommerce par exemple.

Dr BATOUMA NARKOY Master Informatique : BDA

59

Les Bases orientées colonnes

- Les bases orientées colonnes peuvent être vues comme une **extension des tables relationnelles**. On y retrouve le principe des tables et des colonnes.
- La principale différence avec le modèle relationnel est que les colonnes (schéma des tables) **ne sont pas identiques à chaque ligne**. Elles utilisent souvent des clés (clés primaires et clés de tri) pour requêter efficacement dessus.
- Cela va optimiser l'espace de stockage car l'historisation peut être effectuée **sur une seule ligne**.
- L'autre particularité vient aussi des valeurs possibles : dans une colonne, on peut avoir des structures plus élaborées que simplement des nombres ou des textes. Il est possible d'avoir des **structures d'objets ou des listes** comme valeurs.

Dr BATOUMA NARKOY Master Informatique : BDA

60

Les Bases orientées colonnes

Product ID	Name	Price 1	Price 2	Price 3
1	Liquide vaisselle	date: 01/02/2020 price: 2.42€	date: 12/05/2020 price: 2.48€	
2	Shampooing	date: 08/05/2020 price: 1.56€	date: 12/09/2020 price: 1.12€	date: 19/09/2020 price: 1.56€
3	Fromage blanc	date: 12/05/2020 price: 2.02€		

Dr BATOUMA NARKOY Master Informatique : BDA

61

Les Bases orientées documents

- Dans les bases orientées documents, on considère des collections de documents, où chaque document contient une liste de champs **clé/valeur**.
- Le format utilisé dans les documents est principalement le **JSON** ou le **XML**.
- Chaque document possède sa propre structure, ce qui permet d'être très flexible au sein d'une même collection.
- Notons qu'il est possible d'ajouter des **schémas de données** pour obliger un minimum de structure.

Dr BATOUMA NARKOY Master Informatique : BDA

62

Les Bases orientées documents

<p>Document (id: 5baf47)</p> <pre>{ "nom": "Liquide vaisselle", "images": ["https://...", "https://..."], "specs": { "parfum": "Orange" } }</pre>	<p>Document (id: ea53aa)</p> <pre>{ "nom": "Shampooing", "images": ["https://...", "https://..."], "specs": { "parfum": "Vanille" } }</pre>	<p>Document (id: d710bb)</p> <pre>{ "nom": "Fromage blanc", "images": ["https://...", "https://..."], "specs": { "mat_grasses": "0%" } }</pre>
---	---	--

Dr BATOUMA NARKOY Master Informatique : BDA

63

Les Bases orientées documents

- Il faut les utiliser lorsqu'il n'y pas de relations entre les documents et les collections.
- Gestion de catalogues diversifiés (produits, achats, ventes)
- Base d'utilisateurs et de clients
- Web Analytics
- S'il y a une cohérence assez forte entre les données (colonnes plus ou moins similaires), on préférera utiliser une base NoSQL orientée colonnes.

Dr BATOUMA NARKOY Master Informatique : BDA

64

Les Bases orientées clé-valeur

- Les bases orientées clé/valeur sont utilisées pour le stockage temporaire de données ou pour un accès rapide sans requêtes complexes.
- Chaque donnée est référencée par une clé : c'est à partir de cette référence que l'on y accède. Le principe est très similaire aux systèmes de stockage d'objets où on accède aux objets via une clé.

Clé	Valeur
1	https://adresseweb.com
2	356
3	mail: monmail@gmail.com date: 25/10/2020 13:42:12

Dr BATOUMA NARKOY Master Informatique : BDA

65

Les Bases orientées clé-valeur

- Contrairement aux autres bases NoSQL, celles-ci sont plus faciles à utiliser mais leurs usages sont limités. En effet, il n'est pas possible d'effectuer des requêtes poussées sur ces bases.
- La principale utilisation des bases NoSQL orientées clé/valeur concerne la mise en cache d'informations.
- Panier d'utilisateur (E-Commerce).
- Collecte d'événements (tracking utilisateur).
- Partage d'information pour des applications en équilibrage de charge.

Dr BATOUMA NARKOY Master Informatique : BDA

66

Base de données orientés graphe

- La base de données orientée **Graphe** est une base de données composée de **nœuds** et de **bords**.
 - Les nœuds représentent **la valeur de la donnée**
 - Les bords représentent **les relations** qu'entretiennent les nœuds les uns par rapport aux autres.
- Les bords peuvent avoir une direction comme sur l'exemple ci-dessous. Ce type de données permet d'effectuer une analyse **rapide** des relations qu'entretiennent les données.
- L'exemple le plus célèbre est celui des **réseaux sociaux** : dans ce type de base les profils sont des nœuds et les amitiés sont des bords. Un appel sur un profil permet directement d'identifier les cercles de connaissance d'une personne.

Dr BATOUMA NARKOY Master Informatique : BDA

67

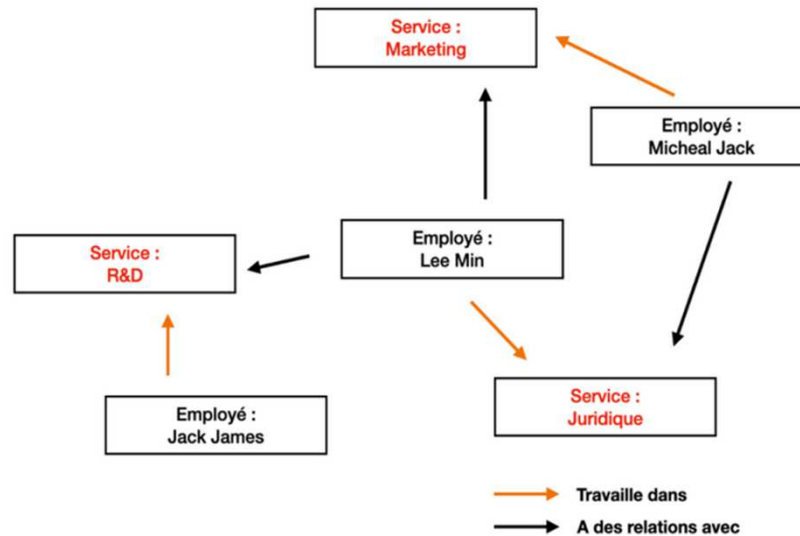
Base de données orientés graphe

- Mais prenons un exemple un peu moins conventionnel mais plus **applicable dans une entreprise**. Disons que nous disposons d'une liste d'employés et de services au sein de l'entreprise.
- De plus, nous savons dans quel service travaillent les employés. Et enfin, avec quels services ils interagissent.
- Nous pouvons créer un **graphe** à partir de ces informations.
- Ici, les nœuds peuvent être soit des services dans l'entreprise, soit des employés.
- Il existe deux types de bords : « Travaille dans » et « A des relations avec ». On remarquera qu'ici les bords sont **orientés** entre les employés et les services :

Dr BATOUMA NARKOY Master Informatique : BDA

68

Base de données orientés graphe



69

Base de données orientés graphe

- Un graphe est beaucoup plus clair qu'une liste de noms.
- La **force** principale de ce système est sa **lisibilité**. Il en est de même pour les ordinateurs. Ils auront plus de **facilité** à comprendre les informations de la database et leurs relations.
- Ce système de BDD a l'avantage d'être compatible avec les algorithmes de **parcours de graphe**. Ces algorithmes permettent d'explorer les données de manière extrêmement **efficace**, c'est pourquoi ce système est particulièrement adapté aux besoins des réseaux sociaux (entre autres).
- Si ta database peut se construire sous la forme d'un graphe comme pour des profils de réseaux sociaux alors l'application de cette base de données est **parfaite** !

Dr BATOUMA NARKOY Master Informatique : BDA

70

Quand utiliser une base NoSQL?

- La **volumétrie est importante** : les bases NoSQL sont adaptées pour stocker et requêter sur des centaines de Go et données. Elle ont été conçues avec l'idée de pouvoir **gérer une forte volumétrie de données**.
- La structure de données est **flexible et peut évoluer** au cours du temps. Si par exemple, une startup lance un produit et ne sais pas encore comment son modèle économique va évoluer au cours des prochains mois, la flexibilité d'une base NoSQL lui permettra **d'adapter son architecture sans trop de difficultés**.
- La base de données peut être **redimensionnée à grande échelle**. Il peut être difficile de mettre à l'échelle horizontalement une base de données SQL, ce qui est plus facile avec du NoSQL, car elles utilisent des mécanismes particuliers (comme les *replicas* ou le *sharding*) afin de gérer de très grandes charges de travail.
- Un des autres avantages des bases NoSQL est qu'elles **ne nécessitent pas autant de modélisation en amont** que les bases SQL. En raison de leur flexibilité,

Dr BATOUMA NARKOY Master Informatique : BDA

71

Bases de Données sur le Web

- La plupart des sites Web s'appuient sur un type de base de données pour organiser et présenter des données aux utilisateurs.
- Chaque fois qu'un visiteur utilise les fonctions de recherche sur ces sites, ses termes de recherche sont transformés en requêtes qu'un serveur de base de données va traiter.
- En règle générale, un middleware (ou intergiciel) connecte le serveur Web à la base de données.
- L'omniprésence des bases de données leur permet d'être utilisées dans presque tous les domaines, des boutiques en ligne au micro-ciblage d'un segment des électeurs dans le cadre d'une campagne politique.

Dr BATOUMA NARKOY Master Informatique : BDA

72