

**Typologie :** Critères pour différencier des types :

- Applications
  - OLTP (online transaction processing)
  - OLAP (online analytical processing)
- Utilisateurs
  - Gestion des données, des transactions, cohérence, intégrité
- Types de données
  - Structurées
  - Non structurées, hétérogènes
- Volumes de données
  - Big Data
- Lieu de stockage
  - Centralisée
  - Distribuée

**Propriétés :**

- Indépendance physique des données (la représentation interne des données est transparente à l'utilisateur)

- Indépendance logique des données (plusieurs vues selon les utilisateurs sur un même modèle logique)
- Description des données (Par l'intermédiaire d'un langage de description de données : LDD)
- Accès aux données (Par l'intermédiaire d'un langage de manipulation LMD ou DML)
- Administration des données
- ...

### **ANSI/SPARC à 3 niveaux**

- externe
- conceptuel
- physique

### **Architecture Client / Serveur**

**ACID :** Une transaction est une unité atomique de mise à jour de la bdd vérifiant 4 propriétés :

- Atomicity
- Consistence : les modifications apportées à la bdd doivent être valides, en accord avec l'ensemble de la bdd et de ses contraintes d'intégrité.
- Isolation : les transactions lancées au même moment ne doivent jamais interférer entre elles.

- Durability : toutes les transactions sont lancées de manière définitive.

## **Modèle conceptuel des données : MCD**

- 1ere étape
- étape essentielle
- conditionne la pertinence, la robustesse et la durabilité de la bdd

## **Exemples :**

- Le modèle entité-relation ou entité-association. Mise en correspondance entité / classes

## **Evolutions :**

- Applications : Moteurs de recherche, réseaux sociaux : Mise à l'échelle des données et des traitements
- Infrastructures : Coût du matériel, architecture distribuée.
- Données : Grands volumes, données hétérogènes, ouvertes et mises en lien.

## **Exemple : Le Cloud Computing**

Le cloud computing, abrégé en cloud (« le Nuage » en français) ou l'informatique en nuage désigne un ensemble de processus qui consiste à utiliser la puissance de calcul et/ou de stockage de serveurs informatiques distants à travers un réseau, généralement Internet. Ces ordinateurs serveurs sont loués à la demande, le plus souvent

par tranche d'utilisation selon des critères techniques (puissance, bande passante...) mais également au forfait. Le cloud computing se caractérise par sa grande souplesse d'utilisation : selon le niveau de compétence de l'utilisateur client, il est possible de gérer soi-même son serveur ou de se contenter d'utiliser des applicatifs distants en mode SaaS<sup>1,2,3</sup>. Selon la définition du National Institute of Standards and Technology (NIST), le cloud computing est l'accès via un réseau de télécommunications, à la demande et en libre-service, à des ressources informatiques partagées configurables<sup>4</sup>. Il s'agit donc d'une délocalisation de l'infrastructure informatique.

### **Exemple :** Le Big Data avec NoSql

Les big data, littéralement les « grosses données », ou mégadonnées (recommandé), parfois appelées données massives, désignent des ensembles de données qui deviennent tellement volumineux qu'ils en deviennent difficiles à travailler avec des outils classiques de gestion de base de données ou de gestion de l'information. L'on parle aussi de datamasse en français par similitude avec la biomasse.

Dans ces nouveaux ordres de grandeur, la capture, le stockage, la recherche, le partage, l'analyse et la visualisation des données doivent être redéfinis. Les perspectives du traitement des big data sont énormes et pour partie encore insoupçonnées ; on évoque souvent de nouvelles possibilités en termes d'exploration de l'information diffusée par les médias, de connaissance et d'évaluation, d'analyse tendancielle et prospective et de gestion des risques

(commerciaux, assuranciers, industriels, naturels) et de phénomènes religieux, culturels, politiques, mais aussi en termes de génomique ou métagénomique, pour la médecine (compréhension du fonctionnement du cerveau, épidémiologie, écoépidémiologie...), la météorologie et l'adaptation aux changements climatiques, la gestion de réseaux énergétiques complexes (via les smartgrids ou un futur « internet de l'énergie »...) l'écologie (fonctionnement et dysfonctionnement des réseaux écologiques, des réseaux trophiques avec le GBIF par exemple), ou encore la sécurité et la lutte contre la criminalité.

**Exemple :** Le Web of Data, LOD (Linked and Open Data) et Web Semantique

Le Web des données (Linked Data, en anglais) est une initiative du W3C (Consortium World Wide Web) visant à favoriser la publication de données structurées sur le Web, non pas sous la forme de silos de données isolés les uns des autres, mais en les reliant entre elles pour constituer un réseau global d'informations. Il s'appuie sur les standards du Web, tels que HTTP et URI - mais plutôt qu'utiliser ces standards uniquement pour faciliter la navigation par les êtres humains, le Web des données les étend pour partager l'information également entre machines. Cela permet d'interroger automatiquement les données, quels que soient leurs lieux de stockage, et sans avoir à les dupliquer.

Tim Berners-Lee, directeur du W3C, a inventé et défini le terme Linked Data et son synonyme Web of Data au sein d'un ouvrage portant sur l'avenir du Web sémantique. En France, le terme

Web des données est de plus en plus utilisé par la communauté des professionnels du domaine.

Le Web sémantique, ou toile sémantique, est un mouvement collaboratif mené par le World Wide Web Consortium (W3C) qui favorise des méthodes communes pour échanger des données.

Le Web sémantique vise à aider l'émergence de nouvelles connaissances en s'appuyant sur les connaissances déjà présentes sur Internet. Pour y parvenir, le Web sémantique met en œuvre le Web des données qui consiste à lier et structurer l'information sur Internet pour accéder simplement à la connaissance qu'elle contient déjà.

Selon le W3C, « le Web sémantique fournit un Modèle qui permet aux données d'être partagées et réutilisées entre plusieurs applications, entreprises et groupes d'utilisateurs ».

L'expression a été inventée par Tim Berners-Lee, l'inventeur du World Wide Web et directeur du World Wide Web Consortium (« W3C »), qui supervise le développement des technologies communes du Web sémantique. Il définit le Web sémantique comme « un web de données qui peuvent être traitées directement et indirectement par des machines pour aider leurs utilisateurs à créer de nouvelles connaissances ».

Alors que ses détracteurs ont mis en doute sa faisabilité, ses promoteurs font valoir que les recherches dans l'industrie, la biologie et les sciences humaines ont déjà prouvé la validité du concept original. Les chercheurs ont exploré le potentiel sociétal du web sémantique dans l'industrie et le secteur de la santé. L'article original de Tim Berners-Lee en 2001 dans le Scientific American

a décrit une évolution attendue du Web existant vers un Web sémantique, mais cela n'a pas encore eu lieu. En 2006, Tim Berners-Lee et ses collègues ont déclaré : « Cette idée simple... reste largement inexploitée. »

## **Organisation :**

1. C1 : Rappels
2. C2 : Données Relationnelles
3. TD1 : PostgreSQL
4. C3-C4 : Big Data, JSON
5. TD2-TD3 : NoSQL
6. C5-C6 : Web of Data - XML - RDF
7. TD4-TD5 : SPARQL

