

Notion de Bases de données

1. Introduction à la gestion des données

1.1. Notion de données et d'information

a) Définitions

Données : Les données sont enregistrements de n'importe quelle caractéristique de la réalité sur un support physique qui en garantit leur conservation, leur compréhensibilité et leur repérage.

L'information = Données + Contexte d'interprétation. L'information est constituée par des données insérées dans un contexte d'interprétation.

b) *Éléments constituant une information*

Les données sont généralement regroupées selon leur appartenance à un objet du monde réel. L'ensemble d'objets homogènes (cohérents) constitue ce qui est couramment appelé **entité**.

Une entité est décrite par un ensemble **d'attributs** (ou propriétés). Chacun de ces attributs prend ses valeurs dans un ensemble appelé **Domaine**.

Parfois, l'information provient de deux ou plusieurs entités reliées entre elle par ce qui est couramment appelé **association**.

1.2. La persistance

La durée de vie des données doit être supérieure à la durée des programmes qui les manipulent. La persistance peut être obtenue en effectuant un stockage permanent sur le disque.

1.3. L'organisation papier

Les principaux inconvénients de l'organisation en papier sont:

- Classement: un document ne peut être classé que sous une seule référence limitant ainsi les possibilités de recherche.
- Consultation: le document est difficile d'accès et le délai de mise en disposition peut atteindre plusieurs jours. De plus, pendant sa consultation par une personne, le document n'est plus disponible à d'autres personnes.
- Contraintes de volumes et taille des documents
- Manque de sécurité vu la possibilité de destruction ou de détérioration (volontaire ou non), vol, ...

1.4. L'organisation en fichiers

Les principaux inconvénients d'utilisation des fichiers

1. Redondance des données et difficultés à saisir les liens entre données.
2. Difficulté et lourdeur d'accès aux données : En pratique, pour chaque accès, même le plus simple, il faudrait écrire un programme.
3. Problèmes de partage des données : Pas de contrôle de concurrence dans le cas où plusieurs utilisateurs accèdent aux mêmes fichiers.
4. Pas d'indépendance entre les données et les traitements
5. Problèmes de sécurité : Si tout programmeur peut accéder directement aux fichiers, il est impossible de garantir la sécurité et l'intégrité des données.
6. Problèmes de maintenance.

2. Bases de données : notions de base

2.1. Définition d'une BD

Une base de données (BD) représente l'ensemble (structuré, cohérent, intégré, partagé) de données persistantes décrivant les activités d'une ou plusieurs organisations et dont la gestion est assurée par un programme particulier appelé : Système de Gestion de Bases de Données

- **Exemple :**

Organisation : un lycée

Données : les étudiants, les enseignants, les cours, etc.

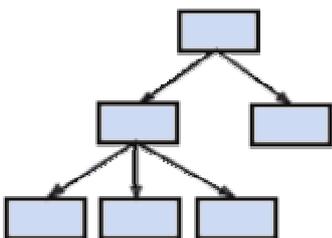
2.2. Intérêts de l'utilisation d'une BD

- **La centralisation :** Les données sont regroupées dans une structure centrale pour éviter les redondances et les incohérences des données.
- **Indépendance Programme/données:** Assurer une Indépendance entre données sur disque et utilisateurs (programmes, humains).
- **Intégration des liaisons entre les données :** les relations entre les données font partie de la base de données et non pas des programmes comme le cas de l'organisation en fichiers.
- **Intégrité des données:** les données doivent être cohérentes entre elles, de plus lorsque des éléments font référence à d'autres, ces derniers doivent être présents.
- **Partage de données (ou concurrence d'accès) :** les données sont utilisées par beaucoup d'utilisateurs. Chacun doit avoir l'impression qu'il est seul mais les données doivent rester cohérentes.

2.3. Les modèles des BD

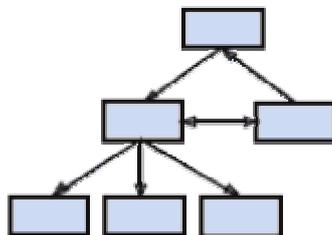
a) Le modèle hiérarchique

- Schéma BD : arborescence
- BD : collection d'enregistrements reliés par des pointeurs avec des liens de type 1-n
- Problèmes : pas d'indépendance, redondance
- Exemple : IMS d'IBM conçu à la fin des années 60 pour le programme Apollo (NASA)



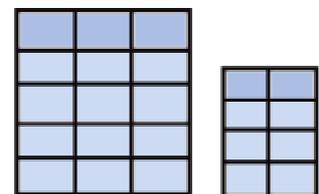
b) Le modèle réseau

- Schéma BD : arborescence
- BD : collection d'enregistrements reliés par des pointeurs avec des liens de type n-n
- Problèmes : pas d'indépendance
- Exemple : IDS2, Total, Adabas



c) Le modèle relationnel

- Schéma BD : Ensemble de relations (tables)
- BD : collection de tables reliées entre elles
- Avantage : indépendance logique / physique
- Exemple : DB2, Oracle, PostgreSQL, Sybase, ...



⇒ C'est le modèle le plus utilisé aujourd'hui