

BASES DES DONNEES

INTRODUCTION

A quoi sert une base de données ? Gérer par exemple une entreprise de transports publics. Un réseau comprend des lignes, des véhicules de différents types et éventuellement des chauffeurs. La BDD nous aidera à savoir par exemple que Marcel est en congé le 30/10 et qu'il assure le 31/10 la ligne H avec le véhicule 56.

A une base de données on pourra poser des questions : dans notre exemple : un véhicule doit-il toujours assurer la même ligne ? Qui a assuré une liaison à une date donnée...

Comment enregistrer et utiliser les informations concernant l'entreprise ?

BESOINS

- ⇒ **Stocker** des gros volumes d'infos
- ⇒ **Partager** des infos
- ⇒ **Gérer l'accès** à ces infos
- ⇒ **Gérer des infos cohérentes et non redondantes.** (non redondantes = optimiser la place)
- ⇒ **Retrouver** l'information **simplement** en posant la bonne question

DEFINITION

BASE DE DONNEES

C'est un ensemble structure de données (prix, numéro d'article...) persistantes, **représentant une partie du monde réel**, qui peut être **partagé** par plusieurs applications (en même temps, possibilité de consulter en même temps), **servant de supports à des applications** informatiques et qui peut être **interrogeable** par le contenu.

Une **BDD doit traduire la connaissance de faits élémentaires** (Marcel est affecté à la ligne H le 24/3/18), de propriétés (un chauffeur a un seul permis de conduire, mais est compétent pour plusieurs véhic), d'évènements (bus 124 supprimé de la circulation)

Une base de données s'appuie sur un système de gestion de base de données (SGBD)

SGBD

C'est un ensemble de **logiciels** qui permet **d'écrire de structurer de mémoriser, manipuler, rechercher de manière efficace des données** spécifiques dans une grande quantité d'informations, stockées sur des mémoires secondaires (disques magnétiques, cloud..) partagées de manière transparente (peut être consultée par plusieurs utilisateurs) par plusieurs utilisateurs = **accéder à des infos en même temps et bcp d'utilisateurs**

Elle assure **la confidentialité** (les usagers n'ont pas à connaître les horaires ou la paie des chauffeurs) et la **sécurité** (définir quelles informations doivent être protégées de quelles personnes. L'utilisateur doit pouvoir consulter un site marchand, les prix, écrire des commentaires, ajouter des articles à son panier par exemple mais sans pouvoir modifier le prix des articles par exemple) **des données dans un environnement multi-utilisateurs.**

Les premiers SGBD datent des années 60 sur des fichiers stockés sur des bandes magnétiques.

On utilise aujourd'hui des SGBDr (Système de Gestion de Base de Données Relationnelles)

Relationnel = sous forme de tableau

OBJECTIFS

Objectif	Description, explications
Indépendance physique des programmes aux données :	Pouvoir modifier les schémas internes sans modifier les schémas conceptuels externes
Indépendance logique des programmes aux données	Garantir l'indépendance entre les différents utilisateurs (la base peut évoluer sans gêner le fonctionnement pour l'utilisateur)
Manipulation des données par des langages non procéduraux	<ul style="list-style-type: none"> ➔ Facilité pour les utilisateurs de manipuler les données ➔ Décrire ce que l'on souhaite sans décrire comment l'obtenir
Administration (centralisée) des données	Outils pour définir et faire évoluer les données
Efficacité d'accès aux données	Optimiser les temps de réponses
Contrôle de la cohérence des données	Eviter la duplication d'infos
Cohérence des données	Vérification des contraintes d'intégrité (=propriété qui doit être vérifiée par exemple pas de notes < 0
Concurrence d'accès (partage de données)	Permettre l'accès simultané aux données tout en conservant l'intégrité de la base (détecter et traiter les éventuels conflits)
Sécurité des données	<ul style="list-style-type: none"> ➔ Protection contre les accès non autorisés pour permettre la confidentialité ➔ Protection en cas de panne (du SGBD ou de la machine) : restauration de la BD

FONCTIONS

Fonction	Intérêt, Description, Définition, Exemple(s) ...
Permettre la description des données	<ul style="list-style-type: none"> ➔ Administrateur de données ➔ Dictionnaire de données : ensemble de schémas et de règles de passage entre schémas associés à une BD pour donner une signification aux données : langage commun.
Permettre la recherche de données	➔ Via un langage d'interrogation : ici le langage SQL
Permettre la mise à jour des données	➔ Insérer, modifier, supprimer...
Transformation des données	
Contrôle de l'intégrité des données	Par exemple : pas d'âge négatif par exemple, dans un banque d'un compte à un autre : pour 100€ qu'il y ait 100 euros enlevé au premier, 100 euros remis sur le second
Concurrence et gestion des transactions	
Contrôle d'intégrité	<ul style="list-style-type: none"> ➔ Conditions qui garantissent la validité de la donnée (pas de prix négatif) ➔ Dépendance entre données (pas 2mêmes numéros étudiants) ➔ Règles d'évolution entre états des données existantes (montant salaire pas diminué)

Schéma	Définition
Schéma Externe	Données perçues par les utilisateurs : description d'une partie de la BD correspondant à la vision d'un programme d'un utilisateur ou un groupe d'utilisateurs
Schéma conceptuel	Niveau central : le plus important. La description des données d'une entreprise en termes de types d'objets et de liens logiques indépendante de toute représentation machine
Schéma interne	Stockage des données, les fichiers, l'accès au fichier. Description des données en termes de représentation physique

LA METHODE MERISE

On a un **décalage entre le monde vivant et le monde artificiel**. Un personnage X se pose une question ou une action qu'il va envoyer à la machine. Le monde artificiel va générer une réponse informatique qui doit être traduite sous une forme compréhensible du monde vivant.

Il existe d'autres méthodes : UML/OMT, SADT...

POURQUOI UNE METHODE D'ANALYSE

- Elle permet **de formaliser une réflexion**
- Garder **une trace compréhensible** de la réflexion grâce à une démarche admise et utilisée par tous, des modèles **admis et utilisés par tous**
- Tendre vers **une réalisation adaptée aux besoins** (ne pas faire une usine à gaz si le but est de remplacer un post-it)
- Fournir des **programmes structurés** et donc maintenables

Merise existe depuis la fin des années 70s. Elle repose sur une méthode conception de système d'information et une double approche : les données et le traitement de ces mêmes données.

Des modèles : outils de représentation du système existant et du système futur

Des niveaux : conceptuel, organisationnel et physique.

Les étapes :

Système d'information manuel → Expression des besoins → Modèle conceptuel → Modèle logique → Modèle physique
→ Système d'information automatisé

DONNEES	QUESTION	TRAITEMENT
Modèle Conceptuel de données	QUOI ?	Modèle conceptuel de Traitement
Modèle Logique des données	QUI OU QUAND ?	Modèle Organisationnel des Traitements
Modèles Physique des Données	COMMENT	Modèle Opérationnel des Traitements

LA DEMARCHE

La méthode repose sur 4 étapes

- Etude préalable
- Etude détaillée
- Réalisation
- Mise en œuvre (formation, documentation, installation et initialisation des données = récupérer les données et les intégrer)

UNE DEMARCHE DE TRAVAIL

- Cerner le problème,
- Définir le domaine de l'étude
- Etablir le dictionnaire d'information
- Etablir les diagrammes de flux (comment l'info circule entre les différentes personnes)
- Etablir les modèles conceptuels
- Etablir les modèles logiques
- Etablir les modèles physiques
- Mise en œuvre

LE MODELE CONCEPTUEL DES DONNEES

Les concepts de base													
Concept	Définitions												
Entités	<ul style="list-style-type: none">➔ Elle est pourvue d'une existence propre➔ Conforme aux choix de gestion de l'entreprise➔ Elle peut être un acteur, ... (par exemple : employé, article..)												
Attributs ou propriétés	<p>Définition : Donnée élémentaire qui se qualifie à l'entité à laquelle elle se rapporte (Attribut par exemple : nom ou prénom)</p> <p>Caractéristiques :</p> <ul style="list-style-type: none">➔ Occurrence : valeur que peut prendre la propriété,➔ Domaine de définition : ensemble des valeurs possibles de l'attribut (note ne peut être que des nombres compris entre 0 et 20)												
Relations ou associations	<p>Définition : Lien sémantique reliant un ensemble d'entités et présentant un intérêt pour l'entreprise (exemple : bus et son conducteur ?)</p> <p><i>Association porteuse : relation qui porte des propriétés</i></p> <table><tr><th colspan="2">Dimension d'associations</th></tr><tr><th>Nom de l'association</th><th>Définition</th></tr><tr><td>Binaire</td><td>lien entre deux entités</td></tr><tr><td>Ternaire</td><td>lien entre 3 entités (entité étudiant, prof, matière)</td></tr><tr><td>n-aire</td><td>lien entre n entités</td></tr><tr><td>réflexive</td><td>lien de l'entité sur elle-même</td></tr></table>	Dimension d'associations		Nom de l'association	Définition	Binaire	lien entre deux entités	Ternaire	lien entre 3 entités (entité étudiant, prof, matière)	n-aire	lien entre n entités	réflexive	lien de l'entité sur elle-même
Dimension d'associations													
Nom de l'association	Définition												
Binaire	lien entre deux entités												
Ternaire	lien entre 3 entités (entité étudiant, prof, matière)												
n-aire	lien entre n entités												
réflexive	lien de l'entité sur elle-même												
Cardinalités :	<p>Définition : quantifie le nombre d'occurrence d'une entité qui participent à une occurrence d'une relation</p> <p><i>Cardinalité minimale :</i> combien d'occurrence au minimum (0 ou 1)</p> <p><i>Cardinalité maximum :</i> combien d'occurrence au maximum (1 ou n)</p>												
Identifiants	<p>Définition : Attribut particulier rajouté de façon arbitraire et qui permet d'identifier de façon unique une occurrence de l'entité (Pour être identifiant, la ou le groupe de propriétés ne peut pas prendre plusieurs fois la même valeur sur l'ensemble des occurrences possibles de l'entité)</p>												

DEMARCHE DE CONSTRUCTION D'UN MCD

- Etablissement du dictionnaire d'informations (recherche de toute l'information à gérer)
- Identification des entités, des attributs et des relations
- Regroupement des propriétés par entité
- Représentation des entités
- Recherche des relations
- Recherche des cardinalités
- Vérification, validation du modèle

CONSTRUCTION DU MCD

Il est établi à partir : du diagramme acteur/flux, des interviews des acteurs impactés par le projet, de l'analyse des documents existants et ... :

- **Le dictionnaire d'informations** : recense toutes les informations utiles au système considéré, est formalisé par un tableau (nom, description, nom dans MCD, type, nom de code), suit toutes les étapes de l'analyse et la mise en place. Il sera la base de la documentation de l'application.
- **Représentation des données**
 - Une donnée est caractérisée par : un nom, une définition, un domaine de définition, une provenance, un mode de calcul (si donnée calculée par exemple il faut : pour un montant TTC, le prix unitaire HT d'un article, le taux de TVA par type d'article, la quantité commandée)
- **Regrouper les attributs par entité** :
 - Un attribut ne peut qualifier qu'une entité et une seule et son nom doit être unique (exemple : Nom_client, Nom_Fournisseur, Date de naissance)
 - Un attribut doit être atomique (exemple : liste_de_notes n'est pas un attribut valable)
- **Représentation des entités** :
 - Première ébauche du modèle conceptuel des données ne faisant apparaître que les entités
- **Recherche des associations** : verbes permettant d'établir des liens logiques
- **Identifiant**
 - Ajouter un identifiant à chaque entité de sorte à ce que chaque enregistrement soit unique
- **Recherche de cardinalités**

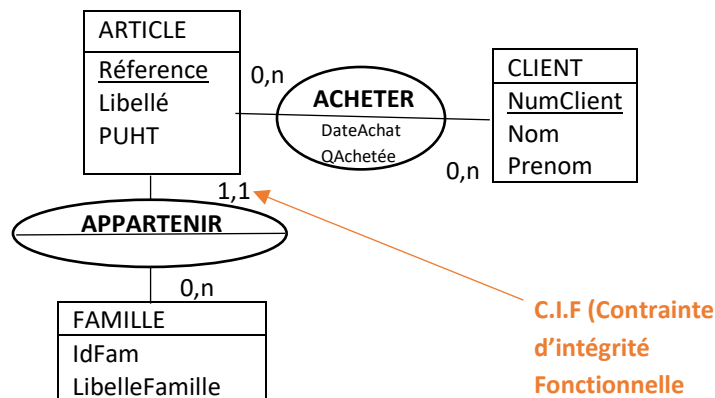
EXEMPLE :

Un magasin vend des articles à des clients à une date donnée. Un article est décrit par son libellé et son PU HT. Un client est défini par son nom, prénom.

Un article appartient à une famille et une seule (par exemple : alimentaire, textile, jouet...)

DICTIONNAIRE DE DONNEES ET REPRESENTATION

Entité / Attribut	
Entité	Article
Entité	Client
Attribut	Date_vente
Attribut	Libellé
Attribut	PUHT
Attribut	Nom
Attribut	Prenom
Entité	Famille
	NumClient
	idFam
	Reference



BASES DE DONNEES : SUITE

APPROCHE PAR NIVEAU DES BDD (RAPPEL)

	Données	Traitement	Niveau
Déroutement du temps et stabilité décroissante ↓	MCD	MCT	Conceptuel
	MLD	MOT	Logique
	MPD	POpT	Physique

SOFTWARE CONSEILLES

Nous ferons usage de différents outils pour effectuer l'analyse (MCD) tel que :

⇒ ANALYSE SI (Gratuit)

⇒ JFreeSoft (Gratuite mais limitée en nombre d'entité dans le modèle relationnel) ⇔ Gère aussi les flux et le MOT.

Ces outils sont capables de gérer : le dictionnaire de données, le MCD, le modèle relationnel et sont capables de créer des scripts SQL permettant de créer des bases de données. Une base de données se constitue de 80% d'études et 20% d'informatique à proprement parler.

Il existe d'autres logiciels payants mais très cher tel que PowerDesigner.

MODELE (RELATIONNEL) LOGIQUE DES DONNEES

A partir du Dico des données est posé le MCD. Ce précédent permet la conception du modèle relationnel : ce qui permettra de passer au niveau logique.

Le niveau logique répond aux questions **qui, où et quand**. Il se situe au niveau logique. C'est une représentation des données, issue de la modélisation conceptuelle puis organisationnelle des données. **Il est exprimé sous forme de relations.**

La modélisation logique des données conduira aux opérations suivantes :

⇒ On transforme le MCD en modèle relationnel. Puis il sera adapté en fonction des spécificités du langage de définition des données.

Le Modèle logique des données est la même chose que le modèle relationnel. Mais est exprimé dans un formalisme différent.

Il a été défini dans les années 70 pour IBM

Un modèle est dit relationnel dans la mesure où il permet de parcourir la structure des données empruntant des chemins non prédéfinis (possibilités de croiser des tables).

A ce type de modèle son associés les notions suivantes : domaine, relation, attribut, tuple (ou n-uplet ⇔ occurrence : une ligne de la table).

MCD	MR
Entité	Relation
Attribut	Attribut
Occurrence	n-uplet ou tuple
Identifiant	clé

Domaine : ensemble de valeurs ayant une signification pour l'utilisateur (ex : le domaine des noms, âges, entiers...)

⇒ La relation est définie comme un tableau de données.

⇒ Les colonnes sont les attributs

⇒ Les lignes des tuples (=occurrence)

⇒ Un ou plusieurs attributs permettent d'identifier de façon unique chaque tuple de la table : la clé primaire de la relation

⇒ La clé primaire est dite simple si elle est composée d'un seul attribut. Dans le cas contraire : c'est une clé composée

Clé primaire	n°	Attribut1	Attribut2
	1	OccurrenceA1	OccurrenceA2
	2	OccurrenceB1	OccurrenceB2

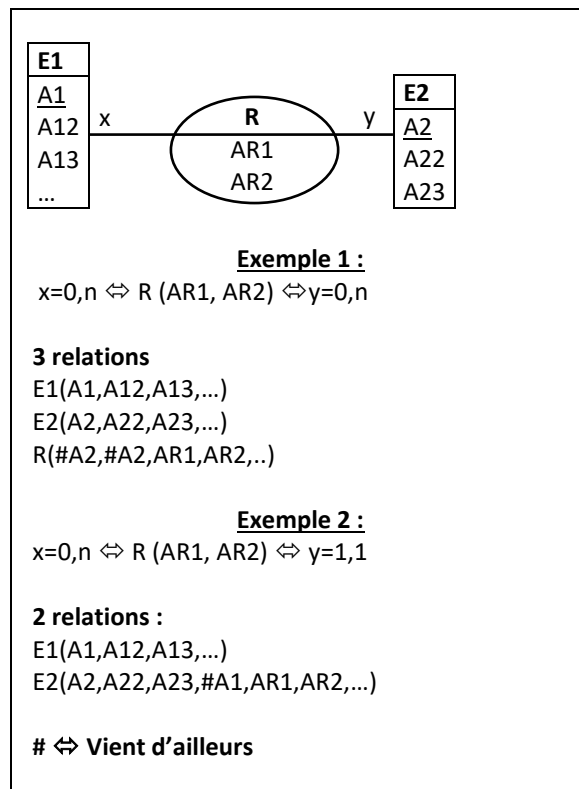
Clé primaire simple : clé composée d'un seul attribut

Clé primaire composée : clé composée de plusieurs attributs

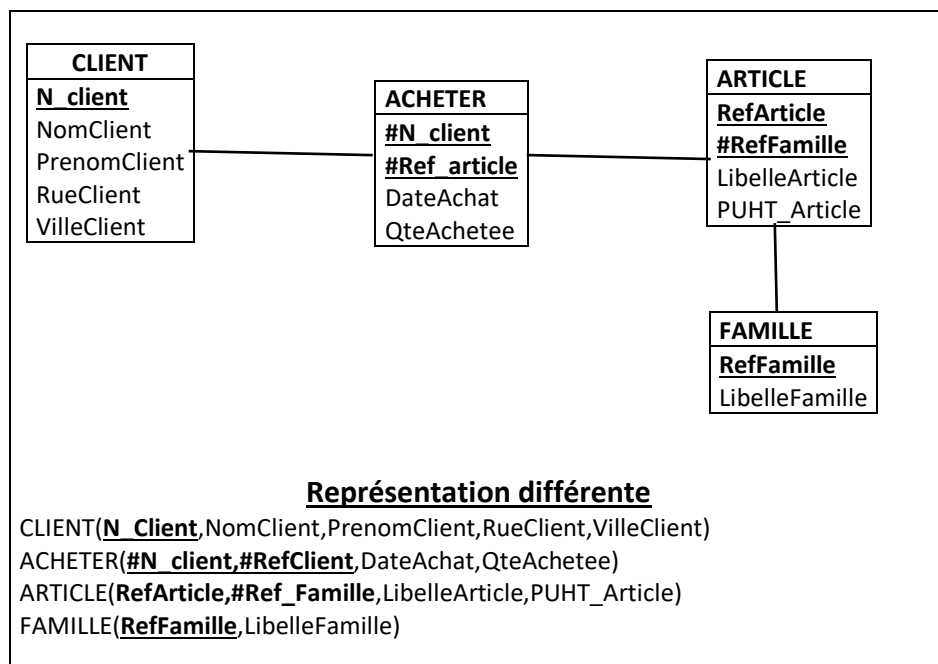
Chaque entité du MCD devient une relation du MR. Mais chaque relation du MCD ne devient pas forcément une relation du MR.

REGLES DE PASSAGE DU MCR

Quand on a une entité reliée par une relation et que de chaque coté ce sont des cardinalités autre qu'une cardinalité 1,1 (CIF). La CIF aspire la relation.



REPRESENTATION : EXEMPLE



\Leftrightarrow Import d'une clé étrangère

Chaque relation ici deviendra ainsi une table

Les clés étrangères permettent de faire de l'algèbre relationnel (= croiser les informations des différentes tables)

Pour une relation réflexive, on importe l'id en tant que clé étrangère TABLE(Id, AH1, AH2, #Id1). #Id1 se rapport à Id.