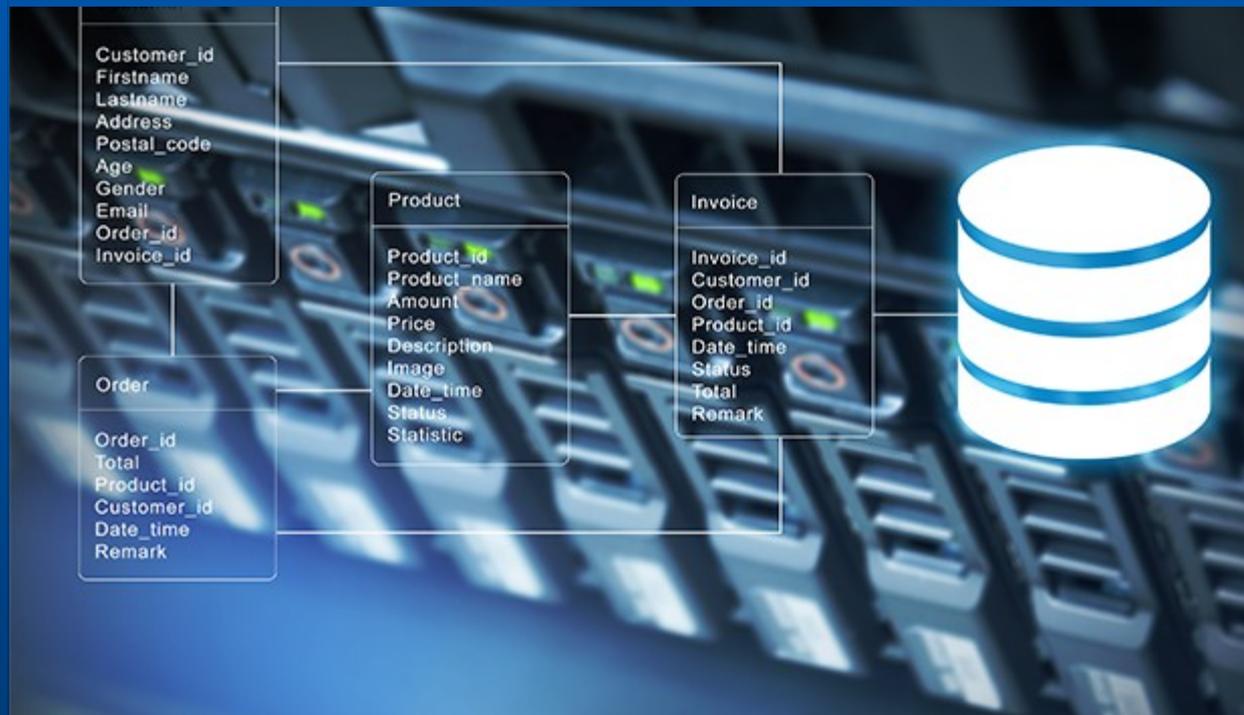


Bases de Données

Numérique & Science Informatique



Bases de Données

Numérique & Science Informatique



Pourquoi stocker des données ?

Conserver l'information...

et y accéder facilement !

Bases de Données

Numérique & Science Informatique



Conserver l'information

Exemple 1 : Magasin de chocolats

- Inventaire complet (détails des stocks disponibles)
- Transactions effectuées
- etc.

Exemple 2 : Site Internet

- Liste des utilisateurs (nom, email, privilèges, etc.)
- Calendrier des évènements
- Commentaires
- etc.

Bases de Données

Numérique & Science Informatique



Accéder à l'information

Exemple 1 : Magasin de chocolats

- Top 10 des produits qui plaisent le plus
- Volume de ventes
- etc.

Exemple 2 : Site Internet

- Top 10 des utilisateurs les plus actifs
- Prochains anniversaires à souhaiter
- Utilisateurs nés en juillet, dont le nom commence par A
- etc.

Bases de Données

Numérique & Science Informatique



Une méthode simple pour stocker l'information

Écrire dans un fichier :

- L'information est stockée sous forme textuelle
- L'information est enregistrée sur la même machine que l'application

```
try :  
    file = open("nomDuFichier.txt", "w")  
    file.write("Voila le contenu du fichier !")  
    file.close()  
except IOError :  
    ...
```

Un module dédié : logging

Bases de Données

Numérique & Science Informatique



Base de données

Ensemble de données structurées stockées à l'écart de l'application.
Fichiers au format XML, CSV, JSON, ...

| Nom | Prénom | Adresse | Numéro de tél. | No appelé | Date et heure de l'appel | Durée de l'appel (sec) | Tarif (CHF/min) |
|------|--------|----------------------------------|----------------|----------------|--------------------------|------------------------|-----------------|
| Camé | Léon | Petit-Chêne 3 2900 Porrentruy | 0324660010 | 00333246634217 | 10.2.2010 14:32 | 455 | 0.70 |
| Camé | Léon | Petit-Chêne 3 2900 Porrentruy | 0324660010 | 0324711230 | 17.2.2010 18:37 | 62 | 0.20 |
| Camé | Léon | Petit-Chêne 3 2900 Porrentruy | 0324669987 | 0324711230 | 23.2.2010 9:21 | 145 | 0.20 |
| Camé | Léon | Grand-Rue 49 2800 Delémont | 0324221022 | 0214537124 | 12.2.2010 19:01 | 302 | 0.10 |
| ... | ... | ... | ... | ... | ... | ... | ... |

L'information ne doit jamais être dupliquée

Bases de Données

Numérique & Science Informatique



Limites de la méthode précédente

Accès et analyse des informations :

- Comment déterminer qui se connecte le plus souvent ?
- ou quels sont les utilisateurs qui ne se connectent que le WE ?

Maintient de la cohérence des informations :

- Comment gérer les doublons ?
- ou les mises à jour en cascades ?

Bases de Données

Numérique & Science Informatique



Une solution...

Un système de gestion de base de données

- stockage, consultation et modification sur une base de données
- offre un langage pour communiquer avec cette base de données
- mise en relation des données



Bases de Données

Numérique & Science Informatique



Propriétés ACID

Le SGBD vérifier les propriétés :

- A** Atomicité : une transaction s'effectue totalement ou pas (rollback en cas d'échec)
- C** Cohérence : une transaction doit respecter TOUTES les contraintes d'intégrités (validité des données)
- I** Isolation : une transaction est indépendante des autres transactions (ie : problèmes des accès concurrents)
- D** Durabilité : une transaction enregistrée ne doit pas être altérée (eg : reprise sur panne)

Bases de Données

Numérique & Science Informatique



Propriétés BASE

Les propriétés ACID ne sont pas applicables dans un contexte distribué. Les propriétés BASE ont alors été proposées pour caractériser les bases NoSQL* :

- BA** Basically Available: le système garantit un taux de disponibilité de la donnée
- S** Soft state: la base NoSQL n'a pas à être cohérente à tout instant
- E** Eventually consistent : à terme, la base atteindra un état cohérent

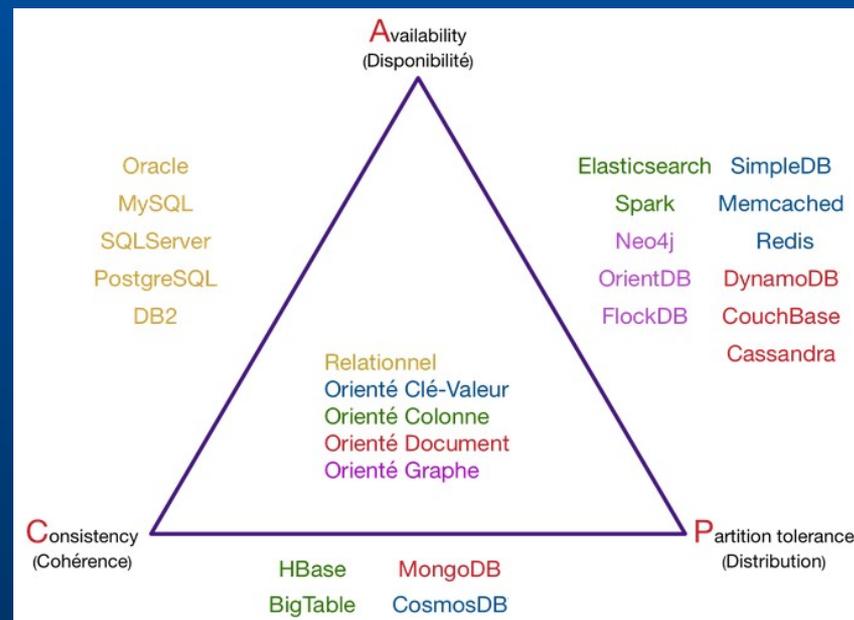
Bases de Données

Numérique & Science Informatique



Théorème de CAP*

Dans toute base de données, vous ne pouvez respecter au plus que 2 propriétés parmi la cohérence, la disponibilité et la distribution.



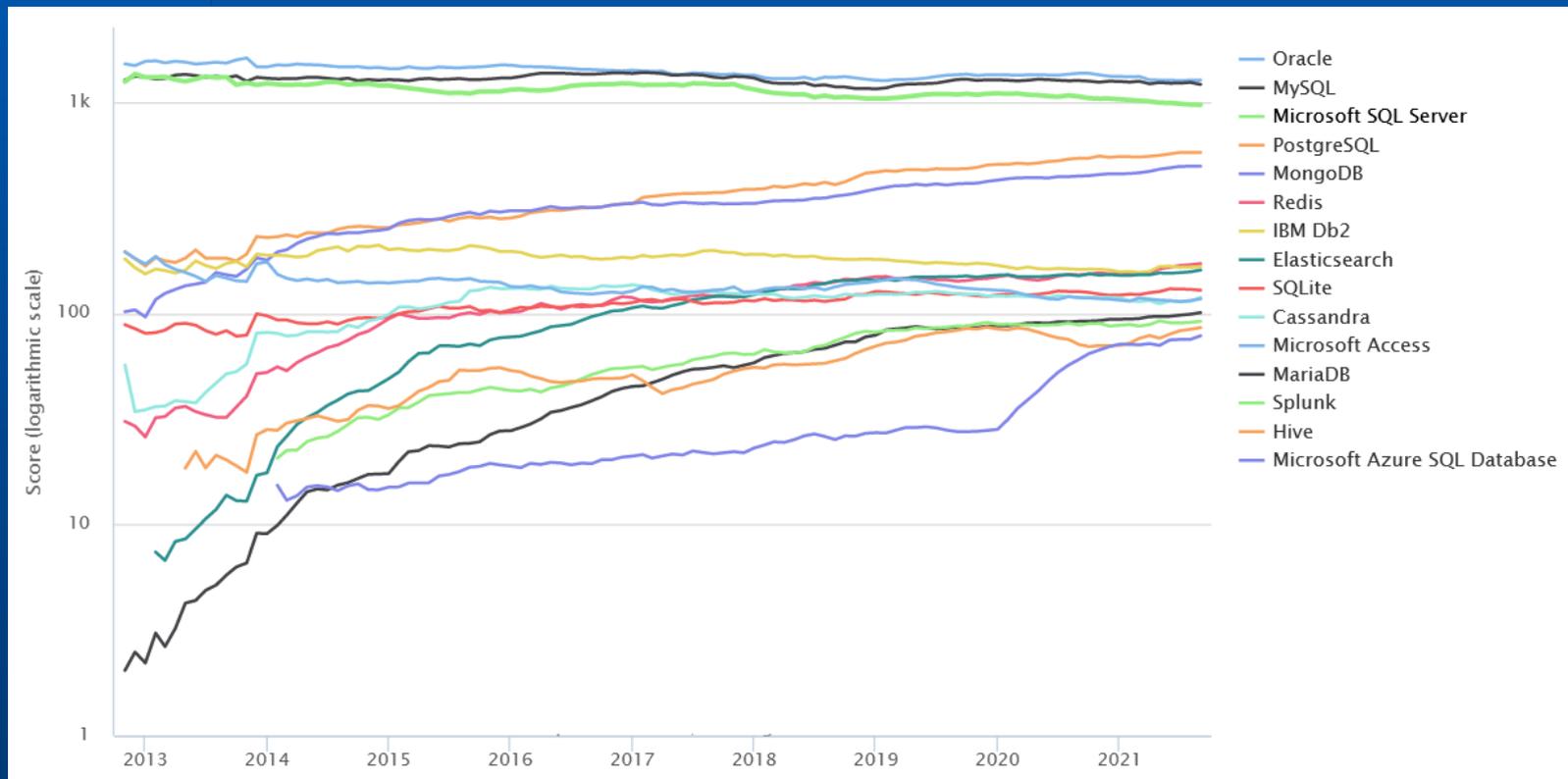
11 * Théorème de Brewer (Eric A. Brewer, 2000)

Bases de Données

Numérique & Science Informatique



DB engines ranking



DB-engines.org

Bases de Données

Numérique & Science Informatique



Moteur de base de données

composant logiciel central qui manipule le contenu des bases de données

Technologies :

- **ISAM*** : les données stockées sont associées à une structure de données en arbre.
- **SQL**** : langage informatique pour formuler des opérations sur les bases de données.

*Indexed Sequential Access Method

**Structured Query Language

Bases de Données

Numérique & Science Informatique



Exemples de moteurs et de SGBDR

- | | | |
|--------------|----------------------|--------------------|
| • SQLite | Domaine public | non ACID |
| • InnoDB* | GNU-GPL/propriétaire | transactionnel |
| • MyISAM | GNU-GPL | non transactionnel |
| • MySQL* | GNU-GPL/propriétaire | |
| • MariaDB | GNU-GPL | |
| • PostgreSQL | BSD | |



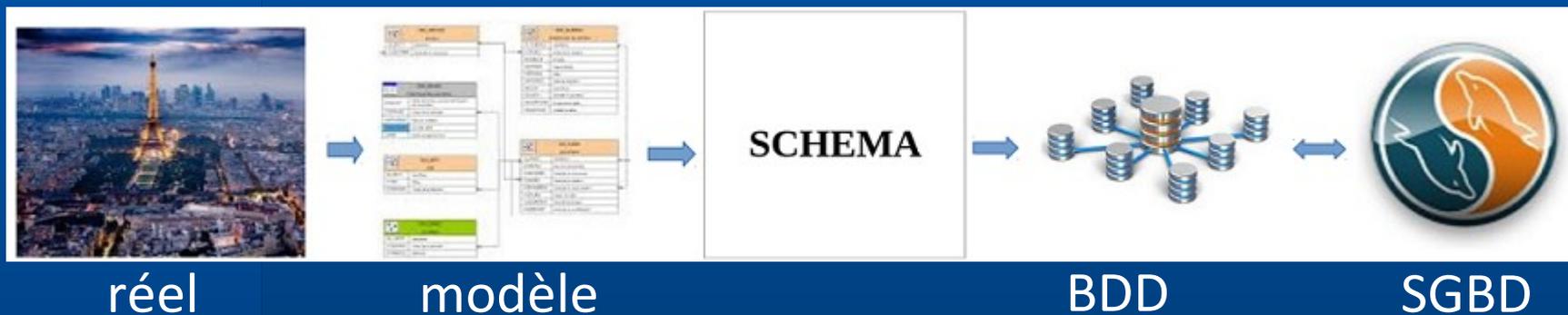
Bases de Données

Numérique & Science Informatique



Base de données relationnelle

Structuration des données pour les mettre en relation



SGBD : logiciel pour manipuler les données (CRUD*)

Bases de Données

Numérique & Science Informatique



Schéma & instance

Le **schéma** : définit la structure des données

L'**instance** : définit leur valeur

Logement (id, nom, capacité, type, lieu)
Activité (idLogement, codeActivité, description)
Client (id, nom, prénom, ville, pays)
Séjour (id, idClient, idLogement, début, fin)

| id | nom | capacité | type | lieu |
|----|------------|----------|-------|--------|
| Pi | U Pinzuttu | 10 | Gîte | Corse |
| Ta | Tabriz | 34 | Hôtel | Alsace |

Bases de Données

Numérique & Science Informatique



Le modèle relationnel

basé sur la notion de relations, représentées par des tables

| idEleve | nom | prenom |
|---------|--------|---------|
| 1 | Bajot | Julien |
| 2 | Balas | Aliocha |
| 3 | Bister | Cyril |
| ... | ... | ... |

| idLog | idEleve | date | details |
|-------|---------|---------------------|---------------------------|
| 1 | 2 | 28/03/2016 17:28:32 | |
| 2 | 1 | 28/03/2016 18:45:01 | |
| 3 | 2 | 29/03/2016 00:05:57 | Erreur d'authentification |
| ... | ... | ... | ... |

Bases de Données

Numérique & Science Informatique



Vocabulaire

Relation

Élève

Nom de l'attribut

| idEleve | nom | prenom |
|---------|--------|---------|
| 1 | Bajot | Julien |
| 2 | Balas | Aliocha |
| 3 | Bister | Cyril |
| ... | ... | ... |

n-uplet (enregistrement)

Attribut (champ)

clé primaire : identifie de manière unique un n-uplet

Bases de Données

Numérique & Science Informatique



Stockage orienté lignes

Le stockage en ligne permet des requêtes relationnelles.

Le stockage en colonne (clef, valeur) distribue les attributs (NoSQL).

Stockage orienté lignes

| id | type | lieu | spec | intérêts |
|---------|-------------------|-----------------|--|------------------------------|
| Nicolas | prof | CNAM | BDD, NoSQL | BZH, Star Wars |
| Régis | | OC | Machine Learning, Dev | escalade, nouilles chinoises |
| Luc | resp formation OC | OC | formation, audiovisuel | |
| Céline | prof | CentraleSupélec | Ontologie, logique formelle, visualisation | |

Stockage orienté colonnes

| id | type | id | lieu | id | spec | id | intérêts |
|---------|-------------------|---------|------------------|---------|------------------|---------|--------------------|
| Nicolas | prof | Céline | Centrale Supélec | Nicolas | BDD | Nicolas | BZH |
| Céline | prof | Nicolas | CNAM | Nicolas | NoSQL | Nicolas | Star Wars |
| Luc | resp formation OC | Régis | OC | Régis | Machine Learning | Régis | escalade |
| | | Luc | OC | Régis | Dev | Régis | nouilles chinoises |
| | | | | Luc | formation | | |
| | | | | Luc | audiovisuel | | |
| | | | | Céline | Ontologie | | |
| | | | | Céline | logique formelle | | |
| | | | | Céline | visualisation | | |

Bases de Données

Numérique & Science Informatique



Le modèle conceptuel des données (MCD)

Description formelle des données du système d'information.



* clef primaire
clef étrangère

Bases de Données

Numérique & Science Informatique



Notation UML

Les cardinalités sont échangées



Bases de Données

Numérique & Science Informatique



Le modèle logique des données (MLD)

Description de la structure du système d'information.

- capteurs_i10n(#IDcapteur, langue, nom)
- capteurs(IDcapteur, unité, min, max)
- mesures(#IDcapteur, date, valeur)

Les clés sont soulignées et les clés étrangères précédées par #.

Bases de Données

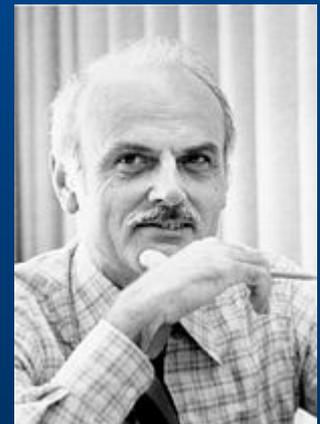
Numérique & Science Informatique



Implémentation du modèle relationnel

Comment passer d'un concept à une solution exécutable ?

- Langage **SQL** (Structured Query Language)
- créé dans les années 90, suite aux travaux d'Edgar « Ted » **Codd**
- adopté en tant que standard en 1987 (**norme** ISO)



Bases de Données

Numérique & Science Informatique



MPD* : implémentation SQL

| Capteurs |
|--------------------|
| IDcapteur : entier |
| Unité : chaîne |
| Min : entier |
| Max : entier |
| primary(IDcapteur) |

```
create table `capteurs` (  
  _ID          int unsigned auto_increment,  
  _unit        varchar(5) not null,  
  _min         int unsigned default '0' not null,  
  _max         int unsigned default '1024' not null,  
  
  check (_unit <> ''),  
  primary key (_ID)  
) ENGINE = MyISAM COMMENT = "Table capteurs";
```

Contraintes :

- Clé primaire, clé unique
- Valeur non nulle
- ...

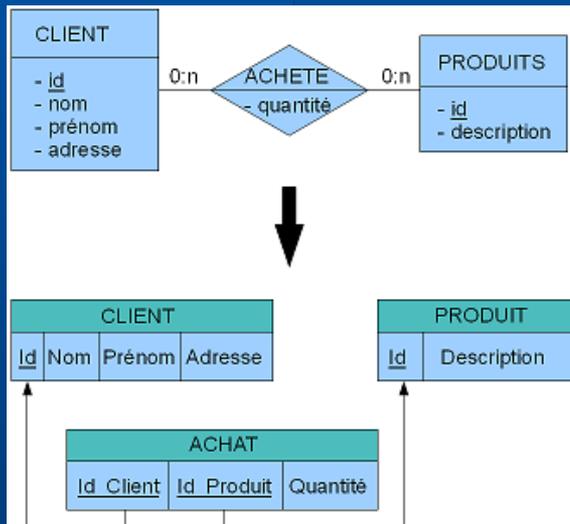
Bases de Données

Numérique & Science Informatique

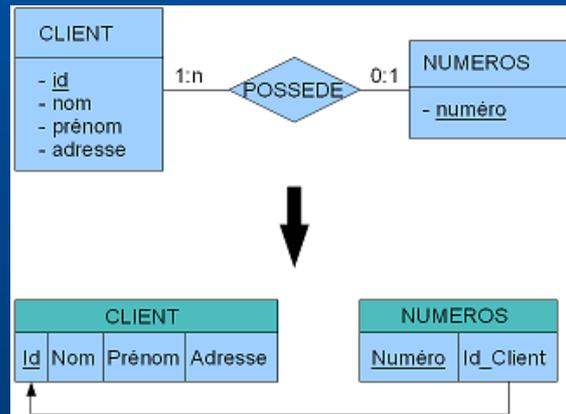


MCD → Modèle relationnel

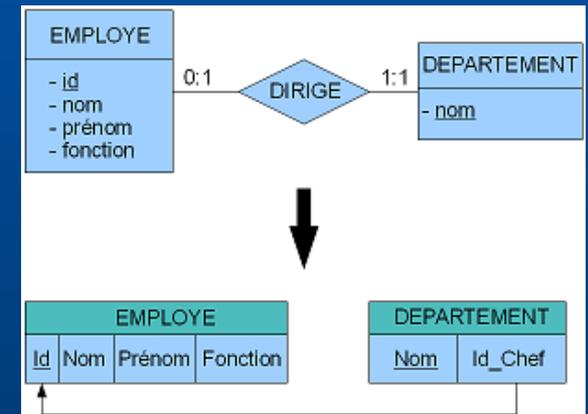
cardinalité **n..n**



cardinalité **1..n**



cardinalité **1..1**



Bases de Données

Numérique & Science Informatique



Formes Normales

La normalisation permet de valider la modélisation des données :

- éviter les redondances
- éviter les incohérences
- limiter le nombre et la durée des mises à jour

=> gain en espace, perte en tps d'exécution

il existe 8 FN :

- 1FN et 2 FN sont nécessaires pour avoir un modèle relationnel juste
- l'usage est de modéliser au moins en 3FN
- les FN sont construites à partir des **dépendances fonctionnelles**

Bases de Données

Numérique & Science Informatique



Dépendance Fonctionnelle

Soit une relation R
un ensemble d'attributs A
un ensemble d'attributs B

quand 2 n-uplets coïncident sur leurs attributs A
et qu'ils coïncident sur leurs attributs B

alors il y a dépendance fonctionnelle entre A et B
on note $A \rightarrow B$

Bases de Données

Numérique & Science Informatique



Normalisation

1FN : attributs représentés par des valeurs atomiques

2FN : tout attribut non clé est en DF minimale avec la clé

ENSEIGNEMENT (IDPROF, CODE_MATIERE, NOM_PROF, VOLUME_HORAIRE)



ENSEIGNEMENT(IDPROF, CODE_MATIERE, VOLUME_HORAIRE)
ENSEIGNANT(IDPROF, NOM_PROF)

3FN : pas de DF Transitive entre attributs non clef

ENSEIGNANT(IDPROF, NOM_PROF, CATEGORIE, CLASSE, SALAIRE)



ENSEIGNANT(IDPROF, NOM_PROF, CATEGORIE, CLASSE)
SALAIRE(CATEGORIE, CLASSE, SALAIRE)

Bases de Données

Numérique & Science Informatique



Exemple

Normaliser la base de données fournisseurs.

| produit | enseigne | adresse | ville | pays |
|------------|-----------|-----------------|--------|---------|
| téléviseur | Vidéo SA | 13 rue de Passy | Paris | France |
| écran plat | Vidéo SA | 13 rue de Passy | Paris | France |
| téléviseur | HiTek Ltd | 25 Bond Street | London | England |

Bases de Données

Numérique & Science Informatique



Exemple

La table est déjà en 1FN

Mise en 2FN

| produit | enseigne | enseigne | adresse | ville | pays |
|------------|-----------|-----------|-----------------|--------|---------|
| téléviseur | Vidéo SA | Vidéo SA | 13 rue de Passy | Paris | France |
| écran plat | Vidéo SA | HiTek Ltd | 25 Bond Street | London | England |
| téléviseur | HiTek Ltd | | | | |

Mise en 3FN

| produit | enseigne | enseigne | adresse | ville | ville | pays |
|------------|-----------|-----------|-----------------|--------|--------|---------|
| téléviseur | Vidéo SA | Vidéo SA | 13 rue de Passy | Paris | Paris | France |
| écran plat | Vidéo SA | HiTek Ltd | 25 Bond Street | London | London | England |
| téléviseur | HiTek Ltd | | | | | |

Bases de Données

Numérique & Science Informatique



Normalisation

BCFN : Forme normale de BOYCE-CODD

3FN est validée +

les DF élémentaires ont une clé qui détermine un attribut

MATIÈRE(NOM-MATIÈRE)
CLASSE(NUM-CLASSE)
ENSEIGNANT(NOM-ENSEIGNANT)
ENSEIGNEMENT(NOM-ENSEIGNANT, NUM-CLASSE, NOM-MATIÈRE)



MATIÈRE(NOM-MATIÈRE)
CLASSE(NUM-CLASSE)
ENSEIGNANT(NOM-ENSEIGNANT, NOM-MATIÈRE)
ENSEIGNER(NOM-ENSEIGNANT, NUM-CLASSE)

Bases de Données

Numérique & Science Informatique



Normalisation

DKNF : Forme normale Domaine Clef

3FN est validée +

toutes les contraintes sont la conséquence logique
des contraintes de domaines et de clés
qui s'appliquent à la relation

VÉHICULE(CONSTRUCTEUR, MODÈLE, TYPE , PTAC)



VÉHICULE(CONSTRUCTEUR, MODÈLE, PTAC)
TYPE VÉHICULE(TYPE , PTAC MIN)

Bases de Données

Numérique & Science Informatique



Transaction

séquence d'opération de lecture et d'écriture permettant de passer d'un état cohérent à un autre

Une transaction se termine par **commit** ou **rollback**

- commit : valide toutes les mise à jour
- rollback : annule toutes les mise à jour

=> propriétés Atomicité + Durabilité

Bases de Données

Numérique & Science Informatique

