

CI 4 : BASES DE DONNÉES

CHAPITRE 1 – INTRODUCTION AUX BASES DE DONNÉES

1 Présentation

1.1 Exemples de bases de données

Si on considère l'ensemble des données présentes sur un disque dur, semblent aisément gérées par un ordinateur. Le temps pour accéder et ouvrir un fichier semble en effet assez court. Mais qu'en est-il lorsqu'il s'agit de trouver un fichier sur le disque ? Qu'en est-il lorsqu'il s'agit de trouver une information dans chacun des fichiers du disque ?

Lorsqu'il s'agit de faire de nombreuses recherches sur un grand nombre de fichiers, un stockage « à plat » ne permet plus un temps d'accès satisfaisant. Il s'agit alors d'organiser les données sous une autre forme. On parle de base de données.

Les bases de données sont omniprésentes dans l'industrie en général et il s'agit de présenter comment ces données sont traitées.

On peut commencer par recenser des bases de données libres. Musicbrainz ou freedb sont par exemples des bases de données qui recensent des informations sur les disques de musiques (Auteurs, compositeurs, interprètes, titres des albums, titres des chansons, dates de sorties ...). IMDb est une base de données cinématographique. OpenStreetMap met à disposition des internautes des données cartographiques.

La FNAC, ou d'autres sites commerciaux disposent d'une base de données de leurs produits. Ainsi, grâce à un champ de recherche, l'internaute peut interroger la base. Il peut avoir des informations sur la disponibilité d'un produit, le délai de livraison ...

Une organisation des données permet aux utilisateurs d'avoir un accès rapide à tout type d'informations. On appelle requête la demande d'un utilisateur formulée auprès d'une base de données.

1.2 Bases de données à «plat»

Une première solution envisageable pour stocker des données est l'utilisation de bases de données dites plates. Les informations sont par exemple stockées dans un tableau. Prenons par exemple une base de données contenant la liste des aéroports du monde ainsi que diverses informations :

Nom	Ville	Pays	Continent	Type
Biarritz-Anglet-Bayonne Airport	Biarritz/Anglet/Bayonne	France	Europe	medium_airport
Milhaud Heliport	Toulon	France	Europe	heliport
Toulon-Hyères Airport	Toulon/Hyères/Le Palyvestre	France	Europe	medium_airport
Lake Hood Seaplane Base	Anchorage	États-Unis	Amérique du Nord	seaplane_base
Ted Stevens Anchorage International Airport	Anchorage	États-Unis	Amérique du Nord	large_airport
Mandalay International Airport	Mandalay	Myanmar	Asie	large_airport

Plusieurs remarques peuvent déjà être formulées :

- plusieurs informations sont stockées à plusieurs reprises : ainsi, l'information «France» est stockée à multiple reprise. Il en est de même pour le champ Continent ;
- des couples d'informations sont redondants : le couple (Pays, Continent) sera toujours identique pour un pays donné. Ainsi, stocker une seule fois que les États-Unis sont en Amérique du Nord.

Dans le cas d'une telle table, des requêtes simples sont aisées. Ainsi, faire la liste de tous les aéroports français ne pose pas de problème. Faire la liste de tous les héliports français est un peu plus difficile.

Si maintenant une seconde table comprend l'ensemble des fréquences sur lesquelles chacun des aéroports peut communiquer, les tables à plat vont rapidement trouver leur limite.

Enfin, lorsque les bases de données deviennent importantes (on recense plus de 40 000 installations aéroportuaires), l'accès aux données situées dans un tableau peut devenir considérablement lent.

1.3 Système de Gestion de Base de Données – SGBD

Pour stocker les données, on utilise des systèmes de gestion de base de données (SGBD). Le marché des SGBD est dominé par les entreprises Oracle, IBM ou Microsoft. Il existe par ailleurs des solutions libres telles que PostgreSQL ou MySQL.

Une SGBD permettent d'assurer le stockage et l'organisation des informations ainsi que les gestions d'accès par des utilisateurs ayant des droits différents. La quantité de données peut dépasser plusieurs TéraOctets.

1.4 Structure client – serveur

Classiquement, les données ne sont pas stockées sur l'ordinateur de l'utilisateur (appelé client) utilisant la base de données mais sur un serveur, voire même un «nuage» (*cloud computing*).

Pour simplifier, pour faire du (*cloud computing*) les entreprises répartissent les informations sur plusieurs ordinateurs en réseau. Suivant les performances nécessaires ou suivant la quantité de stockage nécessaire, la base de donnée peut donc être répartie sur plusieurs ordinateurs physiques, la limite de répartition pouvant être modifié dynamiquement en fonction du besoin.

2 Structure des base de données : le modèle relationnel

2.1 Tables

Dans une première approche, une base de données est constituée de tables. Une table est elle-même constituée de lignes rassemblant les informations (valeurs) que l'on désire stocker. On appellera entité chacune des lignes de cette table. Lorsque les valeurs d'une ligne ont les même propriétés, on les regroupe par colonnes.

Lors de la conception de la base de données, on définit, pour une table, chacune des colonnes. On peut alors renseigner chacune des lignes.

Base de données des installations aéroportuaires

		Aéroports		
Identifiant	Nom	Ville	iso_country	Type
4077	Biarritz-Anglet-Bayonne Airport	Biarritz / Anglet / Bayonne	FR	medium_airport
43537	Milhaud Heliport	Toulon	FR	heliport
4241	Toulon-Hyères Airport	Toulon/Hyères/Le Palyvestre	FR	medium_airport
21567	Lake Hood Seaplane Base	Anchorage	US	seaplane_base
5388	Ted Stevens Anchorage International Airport	Anchorage	US	large_airport
26727	Mandalay International Airport	Mandalay	MM	large_airport

Pays		
Identifiant	Code	Nom
302 687	FR	France
302 649	MM	Myanmar
302 755	US	United States

Exemple

Remarque

Dans une table il n'y a pas de notion d'ordre a priori. Les données d'une ligne ne peuvent donc pas être désignées par un numéro de ligne.

2.2 Définitions [?]

Attributs – Domaines

On considère donné un ensemble infini \mathcal{A} , dont les éléments sont appelés des attributs, un ensemble D , et une application dom de \mathcal{A} dans les sous-ensembles de D .

Si $A \in \mathcal{A}$, $\text{dom}(A)$ est appelé domaine de A .

Définition

Le type d'aéroport est un attribut. Son domaine est l'ensemble des types d'aéroport.

La ville est un attribut. Son domaine est l'ensemble des villes du monde muni d'un équipement aéroportuaire.

Exemple

Schéma relationnel

Un schéma relationnel est un n -uplet de la forme $S = (A_1, \dots, A_n) \in \mathcal{A}^n$ où les A_i sont distincts deux à deux.

Définition

La table des aéroports est un schéma relationnel. Sous une forme formelle, on pourrait noter :

$$\text{Aeroport} = ((\text{Identifiant}, \mathbb{N}), (\text{Ville}, (...)), (\text{iso_country}, (\text{FR}, \text{US}, \dots)), \dots)$$

Exemple

Exemple

Relation – table

On appelle relation associée à un schéma relationnel (A_1, \dots, A_n) est un ensemble fini de n -uplets de $dom(A_1) \times \dots \times dom(A_n)$.

Définition

On note $R(S)$ la relation R pour signifier qu'elle est associée au schéma relationnel S .

Les éléments de R sont appelés les *valeurs*, ou encore les enregistrements, de la relation et leur nombre est appelé son *cardinal* et est noté $\#R$.

Exemple

Les tables présentées précédemment sont des relations.

Définition

Valeurs Si $e \in R(S)$ et $A \in S$, on note $e.A$ la composante du n -uplet e associée à l'attribut A .

R étant un ensemble, deux valeurs distinctes e diffèrent forcément au moins sur un attribut. Formellement, on a :

$$\forall e, e' \in R(S), \text{ si } e \neq e', \text{ alors } \exists A \in S, e.A \neq e'.A$$

2.3 Notion de clef

Afin de ne pas stocker des doublons dans une base de donnée, on a recours au concept de clef primaire.

Dans certains cas, si on est certain que la valeur d'un attribut sera différent pour chaque ligne de la table, cet attribut peut tenir compte de clef primaire. Dans d'autres cas, si on est persuadé que pour une ligne donnée, la combinaison de n attributs est unique, la combinaison de ces attributs peut constituer une clef primaire.

Enfin, dans certains cas, on définit un attribut de type «identifiant». Lorsqu'on ajoute une ligne dans la table, on s'assure qu'un identifiant unique (un nombre entier par exemple) sera affecté.

3 Algèbre relationnelle

La raison d'être d'une base de données étant de disposer de données, il est nécessaire de disposer d'outils d'interrogation.

Définition

Requêtes – Algèbre relationnelle

On entend par algèbre relationnelle, une collection d'opérations (requêtes=) formelles qui agissent sur des relations et produisent une relation en résultat : $R_3 \rightarrow R_2 Op R_1$.

Ceci signifie que dans l'algèbre relationnelle, le résultat des requêtes effectuées sur les relations (tables) sera toujours une nouvelle relation.

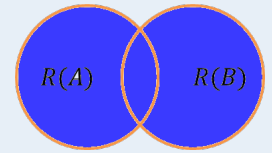
3.1 Opérations sur les ensembles

Définition

Union

L'union de deux relations $R_1(S)$ et $R_2(S)$ est l'ensemble des valeurs comprises dans R_1 ou R_2 .

On peut donc noter la relation $R_3(S)$ définie par : $R_3(S) \leftarrow R_1(S) \cup R_2(S)$

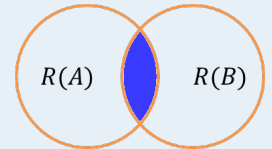


Définition

Intersection

L'intersection de deux relations $R_1(S)$ et $R_2(S)$ est l'ensemble des valeurs comprises dans R_1 et dans R_2 .

On peut donc noter la relation $R_3(S)$ définie par : $R_3(S) \leftarrow R_1(S) \cap R_2(S)$

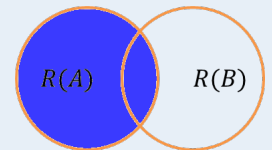


Définition

Différence

La différence de deux relations $R_1(S)$ et $R_2(S)$ est l'ensemble des valeurs comprises dans R_1 et qui ne sont pas comprises dans R_2 .

On peut donc noter la relation $R_3(S)$ définie par : $R_3(S) \leftarrow R_1(S) - R_2(S)$



3.2 Opérations relationnelles

Définition

Projection

Opération notée π au cours de laquelle on sélectionne certaines des colonnes (on élimine donc des attributs).

$$R_2 \leftarrow \pi_{\text{attribut 1, attribut 2, ...}}(R_1)$$

Définition

Sélection

On appelle sélection de R_1 selon $A = a$, et on note $\sigma_{A=a}(R_1)$, la relation obtenue en sélectionnant dans R_1 uniquement les valeurs e telles que $e.A = a$.

$$R_2 \leftarrow \sigma_{\text{attribut=condition}}(R_1)$$

Définition

Renommage d'attribut

Cette opération est utilisée pour des raisons pratiques pour lever une ambiguïté ou pour simplifier le nom d'un attribut de façon temporaire.

Soit $S = (A_1, \dots, A_n)$ un schéma, $i \in [1; n]$ et B un attribut tel que $dom(B) = dom(A_i)$. On note :

$$R_2 \leftarrow \rho_{\text{ancien attribut} \rightarrow \text{nouvel attribut}}$$

Produit cartésien

Soient $R_1(S_1)$ et $R_2(S_2)$ deux relations de schémas disjoints. L'opération produit cartésien est noté \times .

$$R_3 \leftarrow R_2 \times R_1$$

La relation R_3 contient toutes les combinaisons d'association possibles entre les valeurs de R_1 et de R_2 .

Références

- [1] Serge Abiteboul, Benjamin Nguyen, Yannick Le Bras, *Introduction aux Bases de Données Relationnelles Programme de Classes Préparatoires Scientifiques, Première année.*
- [2] Wack et Al., *L'informatique pour tous en classes préparatoires aux grandes écoles, Editions Eyrolles.*
- [3] Christophe Revy, *Concepts des bases de données, Cours de STS IRIS, Lycée Janot de SENS.*
- [4] Patrick Beynet, *Supports de cours de TSI 2, Supports de cours de TSI 2, Lycée Rouvière, Toulon.*