

Base de données et web

La donnée

Qu'est ce qu'une donnée ?

Une donnée est une description élémentaire d'une réalité. Elle est le résultat direct d'une mesure.

Une donnée seule ne permet pas de prendre une décision sur une action à mener.

Ex :

- 35°

Qu'est ce qu'une information ?

Une information est un ensemble de données à laquelle un sens et une interprétation ont été données.

L'information permet de faire des interprétations et/ou de définir des actions à mener.

Ex :

- 35°
- Paris



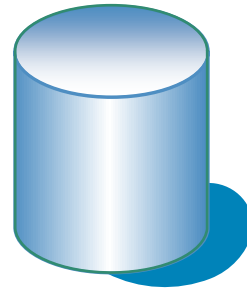
Je connais
maintenant la
température à Paris

Comment transformer la donnée en information ?

Afin de pouvoir exploiter les données et pouvoir en tirer des informations, il faut pouvoir les organiser et les exploiter.

On utilise une base de donnée pour stocker toutes ces données.

Donnée 1
Donnée 2
Donnée 3
Donnée 4
Donnée 5
...



id_matiere	id_voie	id_ue	ue_obligation	id_matiere_ref	taeoin_mere	intitule	intitule_en
1	(NULL)	(NULL)	0	(NULL)	0	Microéconomie	Microeconomics
2	(NULL)	(NULL)	0	(NULL)	0	Macroéconomie	Macroeconomics
4	(NULL)	(NULL)	0	(NULL)	0	Microéconomie 1	Microeconomics 1
5	(NULL)	(NULL)	0	(NULL)	0	Macroéconomie 1	Macroeconomics 1
6	(NULL)	(NULL)	0	(NULL)	0	Théorie des jeux	Game Theory
7	(NULL)	(NULL)	0	(NULL)	0	Commerce international et globalisation	International Trade Theory and glob
9	(NULL)	(NULL)	0	(NULL)	0	Groupe d'éco./socio. théoriques	Theoretical Economics and Sociology
231	(NULL)	(NULL)	0	(NULL)	0	Modèles des marchés imparfaits	Imperfect Market Models
557	(NULL)	(NULL)	0	0	0	Stage de fin d'études MS	Internship
15	(NULL)	(NULL)	0	(NULL)	0	Théorie des institutions	Theory of Institutions
629	(NULL)	(NULL)	0	(NULL)	0	Statistique semi et non paramétrique	Structural Econometrics
630	(NULL)	(NULL)	0	(NULL)	0	Structural Econometrics	Structural Econometrics
16	(NULL)	(NULL)	0	(NULL)	0	Analyse conjoncturelle	Short-Term Macroeconomic Analysis
20	(NULL)	(NULL)	0	(NULL)	0	Development Economics	Development Economics
21	(NULL)	(NULL)	0	(NULL)	0	Séminaire d'économie du développement	Seminar on Development Economics
408	(NULL)	(NULL)	0	(NULL)	0	Topics in insurance economics	Topics in insurance economics
26	(NULL)	(NULL)	0	(NULL)	0	Economic Geography	Economic Geography
27	(NULL)	(NULL)	0	(NULL)	0	Competition and Industrial Policies	Competition and Industrial Policies
227	(NULL)	(NULL)	0	(NULL)	0	Anglais 1	English as a Foreign Language
30	(NULL)	(NULL)	0	(NULL)	0	Health Economics	Health Economics
31	(NULL)	(NULL)	0	(NULL)	0	Economie de l'environnement	Environmental Economics
634	(NULL)	(NULL)	0	(NULL)	0	Estimation non paramétrique	Non Parametric Estimation
633	(NULL)	(NULL)	0	(NULL)	0	Labor economics and Employment Policies	Labor economics and Employment Poli
35	(NULL)	(NULL)	0	(NULL)	0	Macroéconométrie appliquée	Applied macroeconometrics
36	(NULL)	(NULL)	0	(NULL)	0	Thèmes récents en commerce international	Recent topics in international trade
954	(NULL)	(NULL)	0	(NULL)	0	FAITE 2012/2013 Indemnités GREST	(NULL)
955	(NULL)	(NULL)	0	(NULL)	0	Faite 2012/2013 - Concours commun	(NULL)
893	(NULL)	(NULL)	0	(NULL)	0	EDT Gian Li Solheid	(NULL)
814	(NULL)	(NULL)	0	(NULL)	0	Economie de l'innovation et de l'entrep	Economics of innovation and entrepr
38	(NULL)	(NULL)	0	(NULL)	0	Initiation à l'économie	Introduction to Economics

Cependant, comment les exploiter, les insérer, les trier ?

On utilise un SGBD.

Le SGBD

Qu'est ce qu'un SGBD ? (système de gestion de base de données)

Un SGBD est l'interface entre l'homme et la base de donnée.

Il permet :

- D'inscrire des données
- De retrouver des données
- De modifier des données
- De trier des données
- De transformer des données



Transaction informatique

Les types de SGBD

Il existe plusieurs types de SGBD. Les plus connus sont :

- Le modèle hiérarchique
- Le modèle relationnel
- Le modèle objet
- Etc.

La transaction informatique

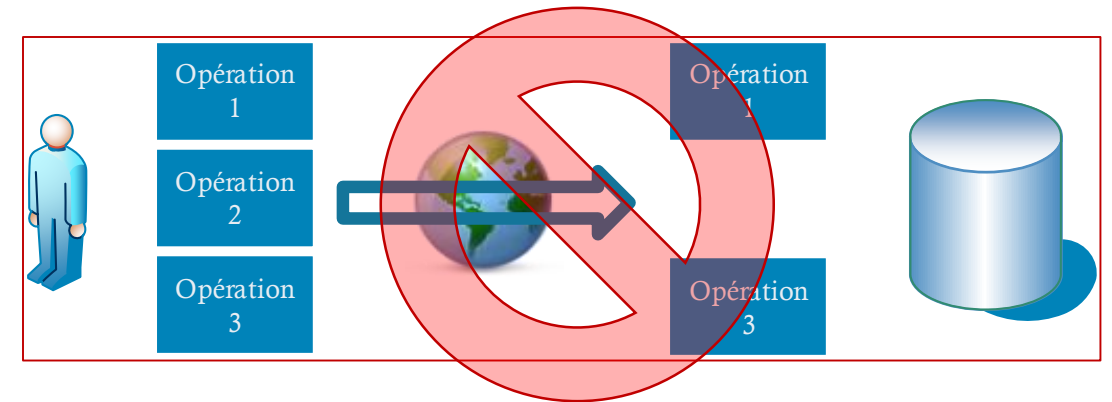
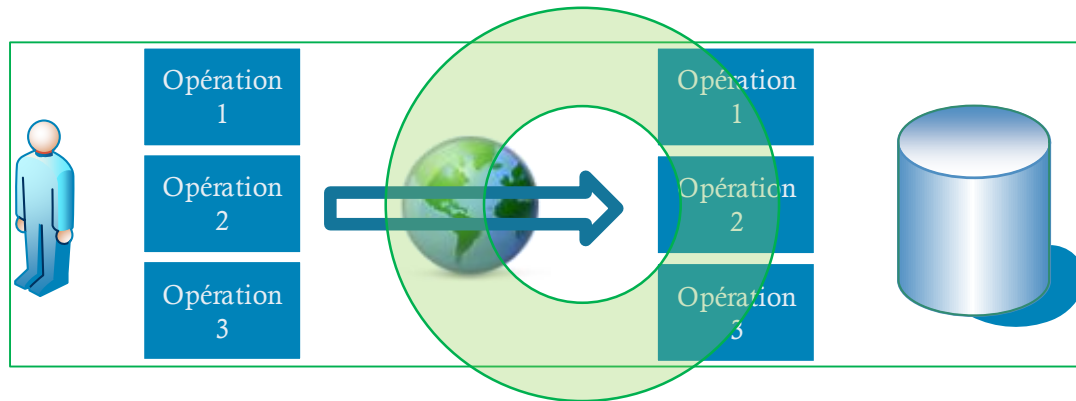
La transaction informatique

- Une transaction informatique est une suite d'opérations visant à altérer l'état d'une base, la faisant passer d'un état A à un état B.
- Une transaction informatique doit suivre à la lettre certaines propriétés appelées ACID.

Atomicité
Cohérence
Isolation
Durabilité

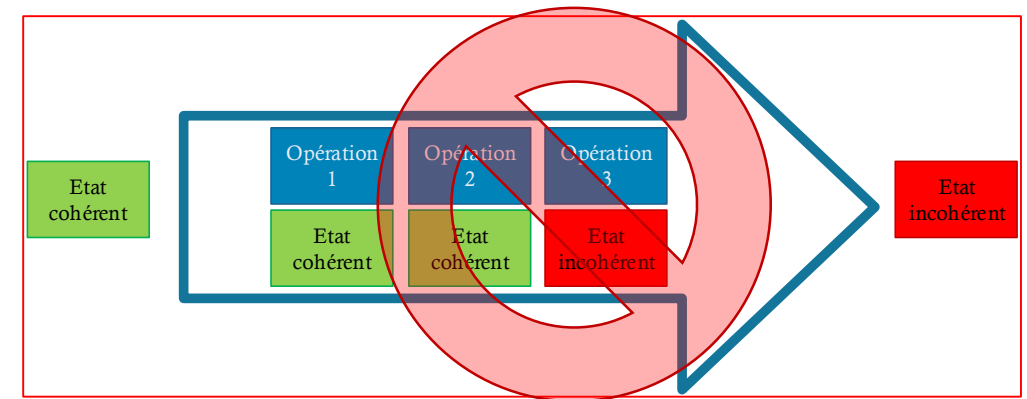
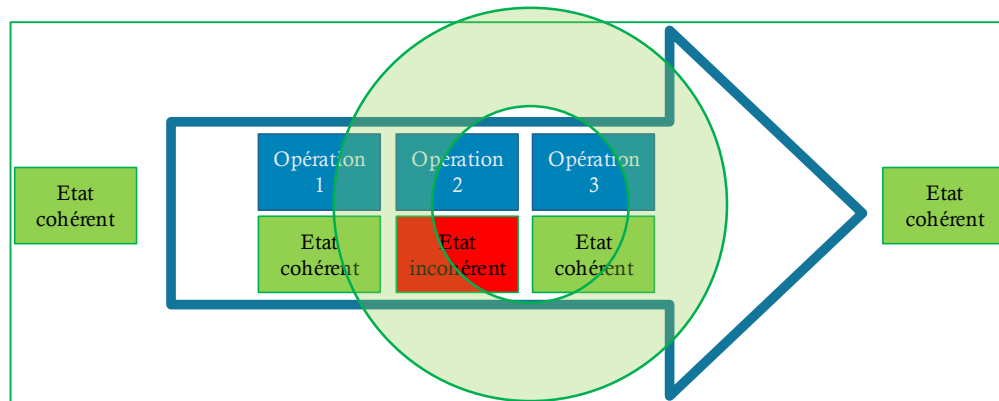
Atomicité d'une transaction informatique

Une suite d'opérations est indivisible, en cas d'échec en cours d'une des opérations, la suite d'opérations doit être complètement annulée quel que soit le nombre d'opérations déjà réussies.



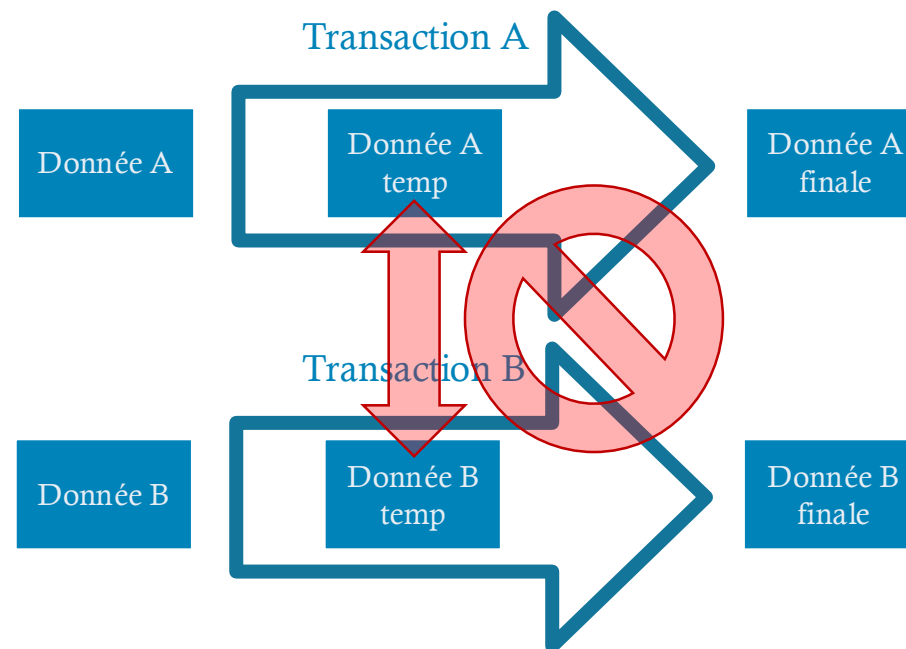
Cohérence d'une transaction informatique

Le contenu d'une base de données à la fin d'une transaction doit être cohérent sans pour autant que chaque opération durant la transaction donne un contenu cohérent. Un contenu final incohérent doit entraîner l'échec et l'annulation de toutes opérations de la transaction.



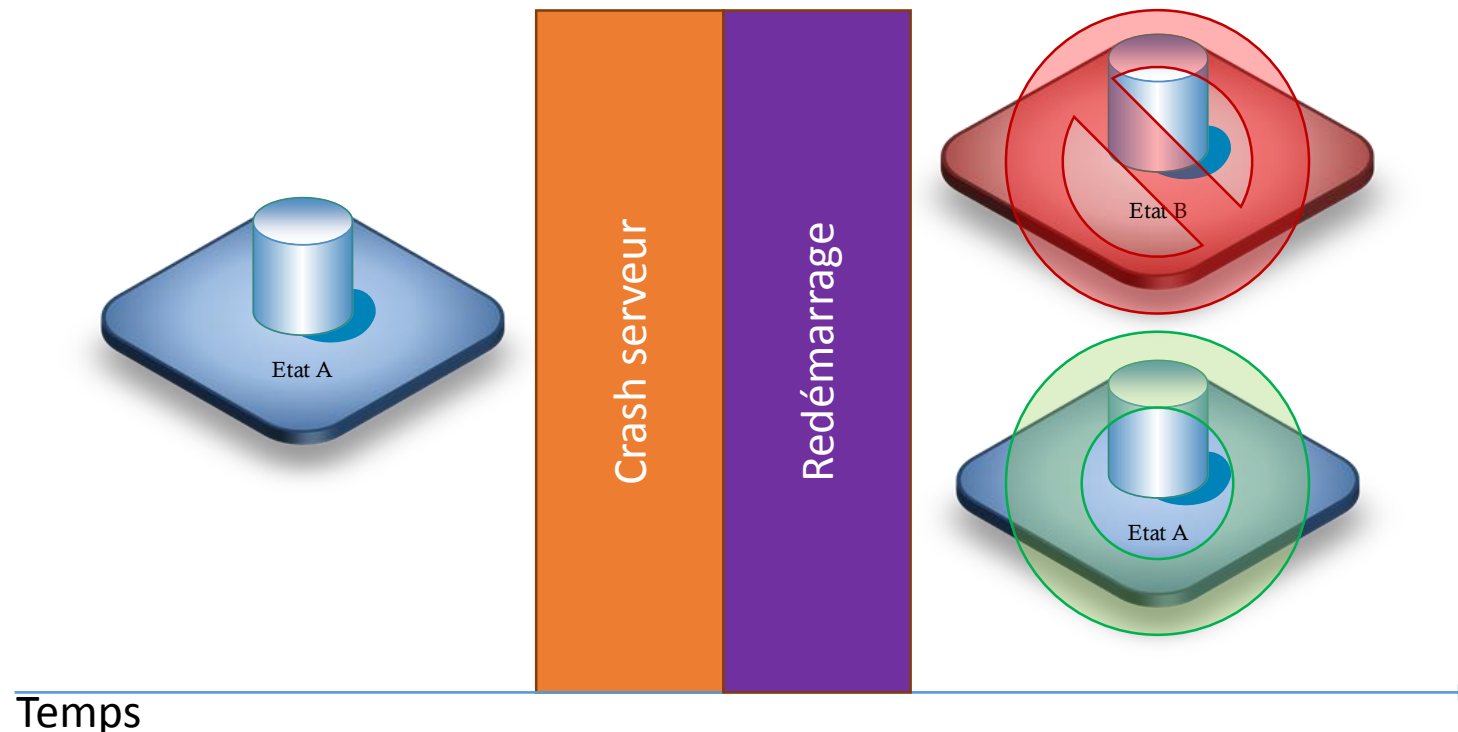
Isolation d'une transaction informatique

Lorsque deux transactions A et B sont exécutées en même temps, les modifications effectuées par A ne sont ni visibles par B, ni modifiables par B tant que la transaction A n'est pas terminée et validée.



Durabilité d'une transaction informatique

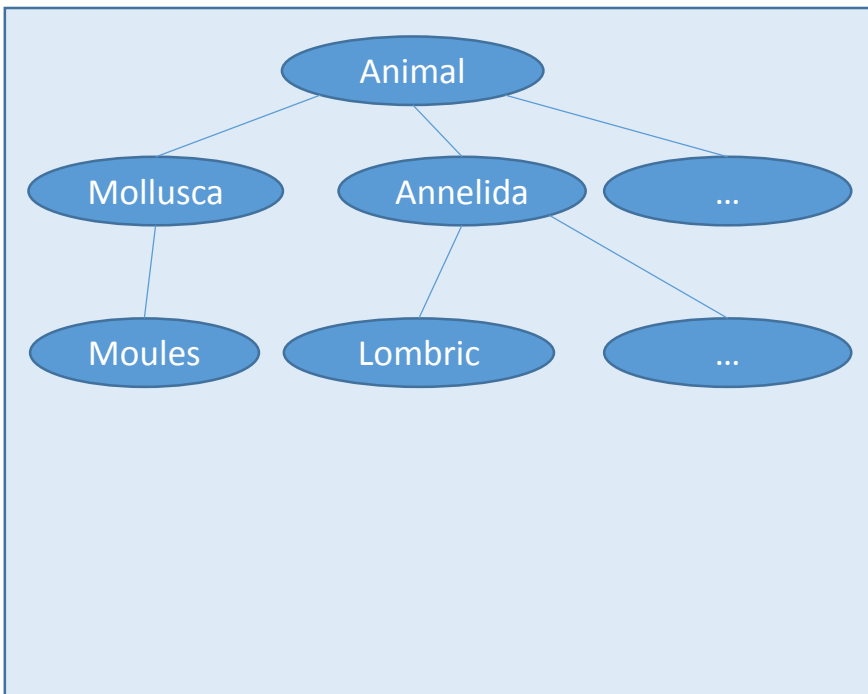
Une fois validé, l'état de la base de données doit être permanent, et aucun incident technique ne doit pouvoir engendrer une annulation des opérations effectuées durant la transaction.



Les modèles de données

Le modèle hiérarchique

- Ce modèle s'inspire de la structure de données en Arbre.
- Chaque entrée n'a qu'un seul parent



- Adapté à la modélisation de nomenclatures ;
- Recherche rapide dans la base.



- Difficulté à décrire une structure complexe ;
- Cardinalité n..n impossible

Le modèle relationnel

- Ce modèle permet d'associer des attributs spécifiques à chaque entité.
- Basé sur l'algèbre relationnelle et la notion de produit cartésien.

Nom	Espèce	Taille moyenne
Moules	Mollusca	12 cm
Lombric	Annelida	7 cm
...		



- Permet d'effectuer facilement des opérations sur la base.
- Recherche rapide dans la base.



- Difficulté à décrire une structure complexe ;
- Ne suit pas l'évolution des SI.

Le modèle objet

- Ce modèle se base sur le modèle relationnel en ajoutant les notions de type objet.
- Ainsi, on peut créer ses propres types, ses fonctions et bénéficier d'héritage.



- Les structures complexes sont plus faciles à mettre en œuvre.



- Problème de scalabilité.
- Pas de réel standard.

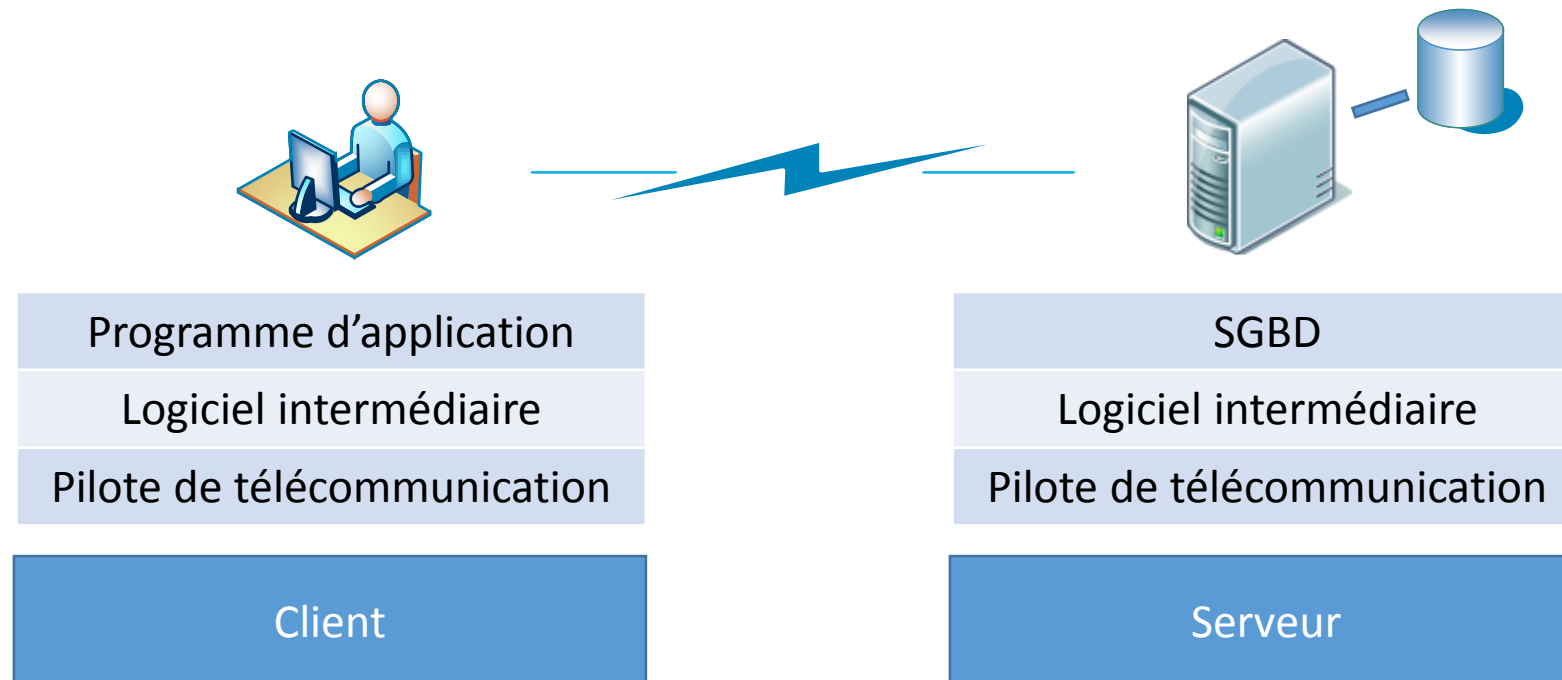
L'architecture d'un SGBD

L'architecture d'un SGBD

- Un SGBD peut être construit de plusieurs manières
 - L'utilisateur peut accéder directement aux données (modèle client – serveur)
 - Les données ne peuvent être accessibles que par un serveur métier (3 tiers)
 - Les données peuvent être réparties à plusieurs endroits différents (modèle distribué)
 - Plusieurs instance peuvent travailler sur la même base de donnée (modèle parallèle)
 - Etc.

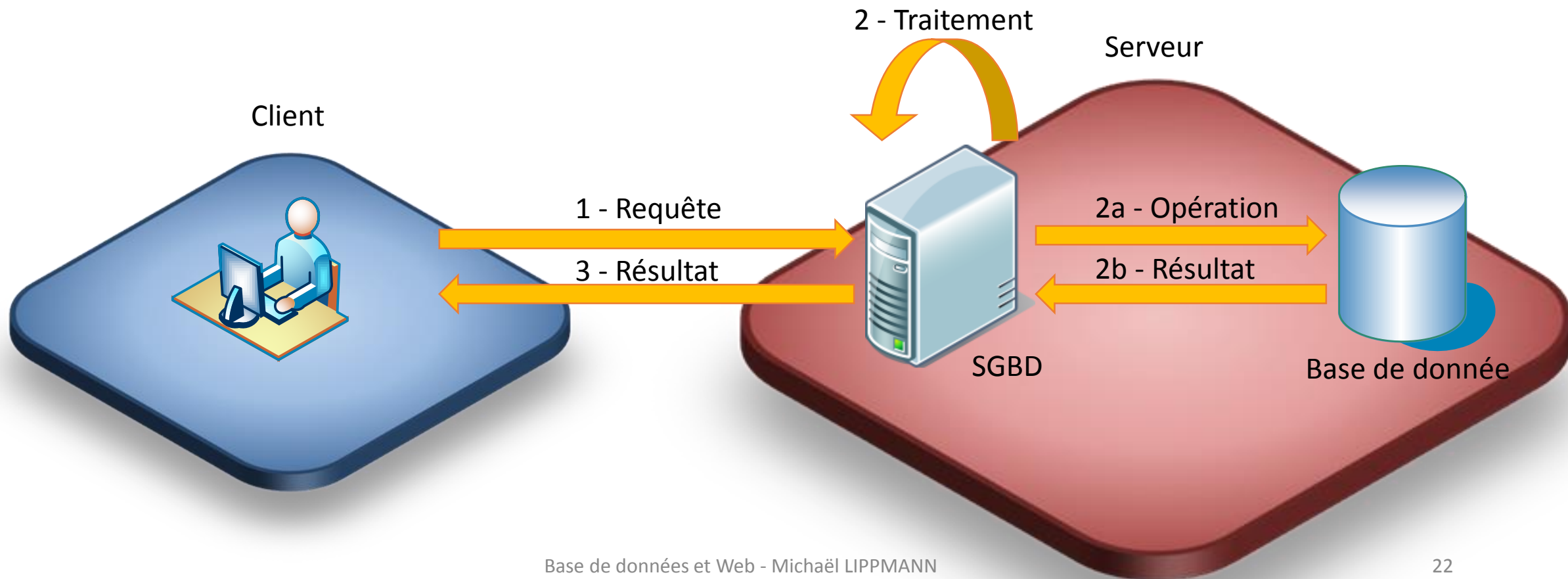
L'architecture d'un SGBD

- Une architecture Client - Serveur



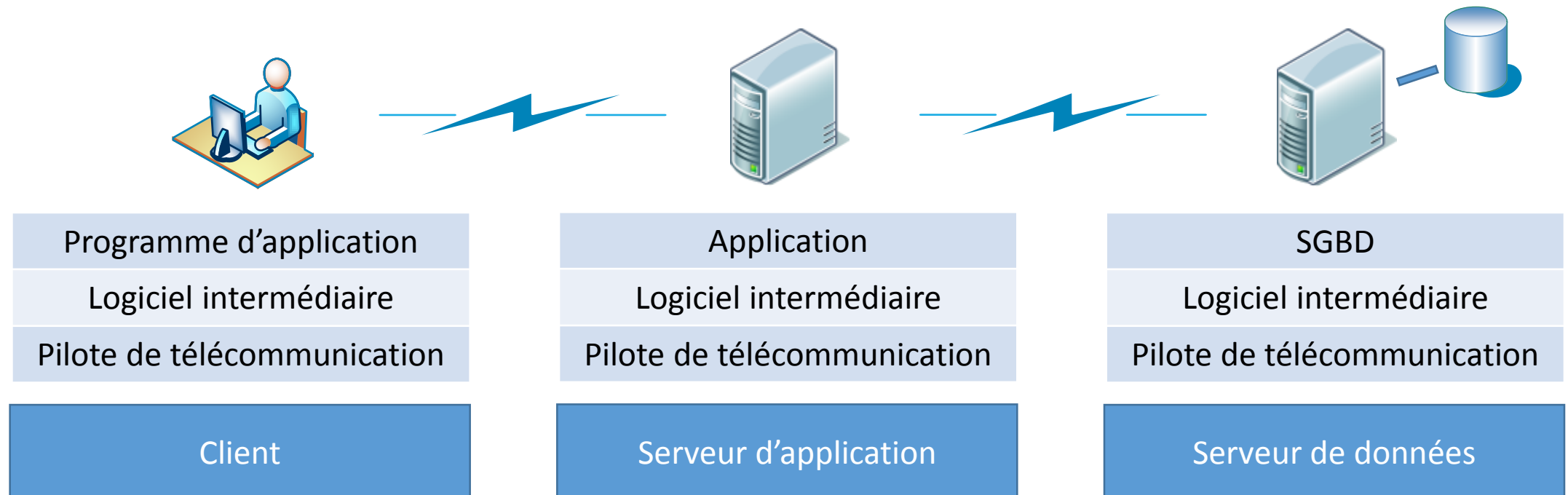
L'architecture d'un SGBD

- Une architecture Client - Serveur



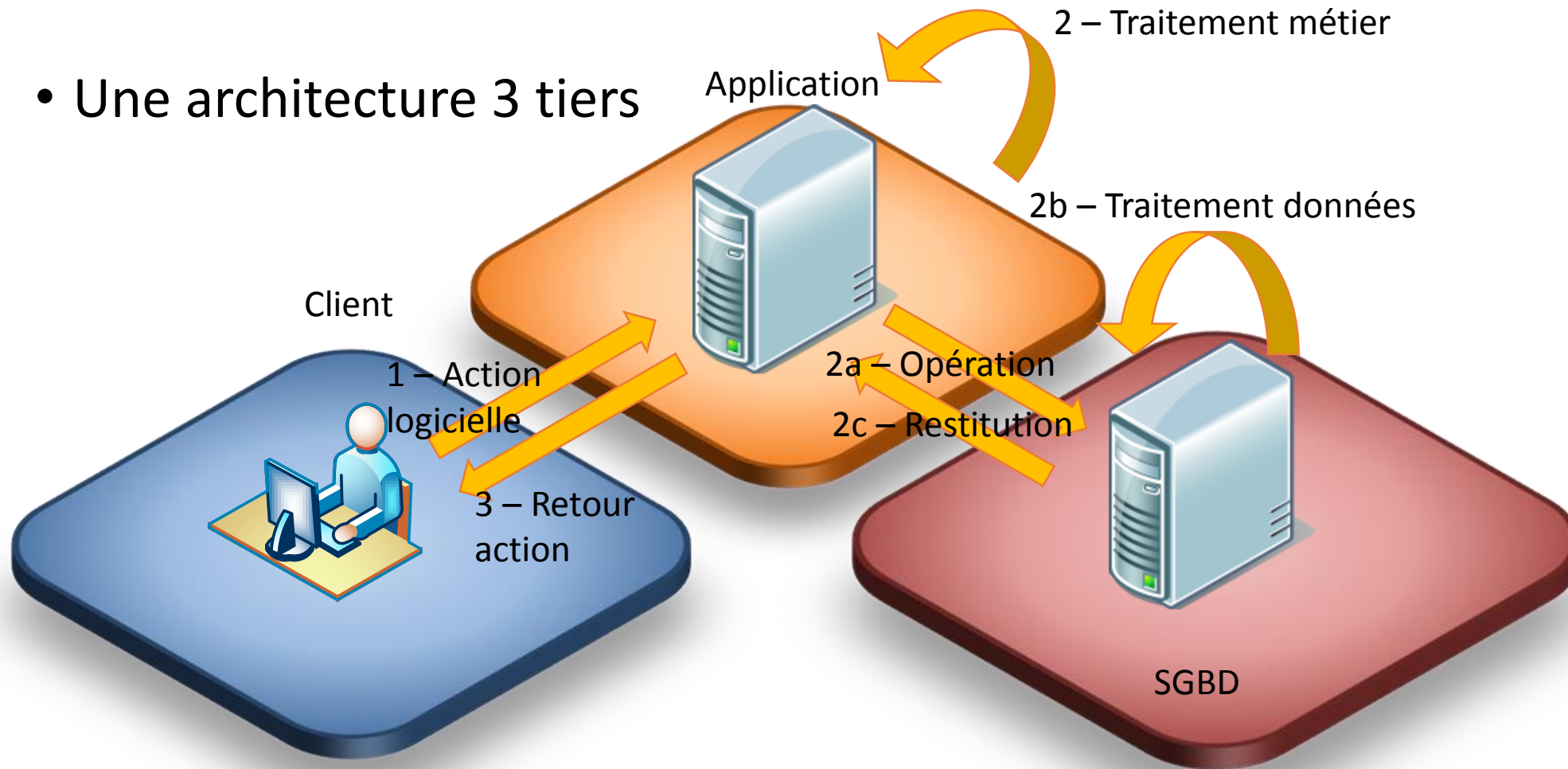
L'architecture d'un SGBD

- Une architecture 3 tiers



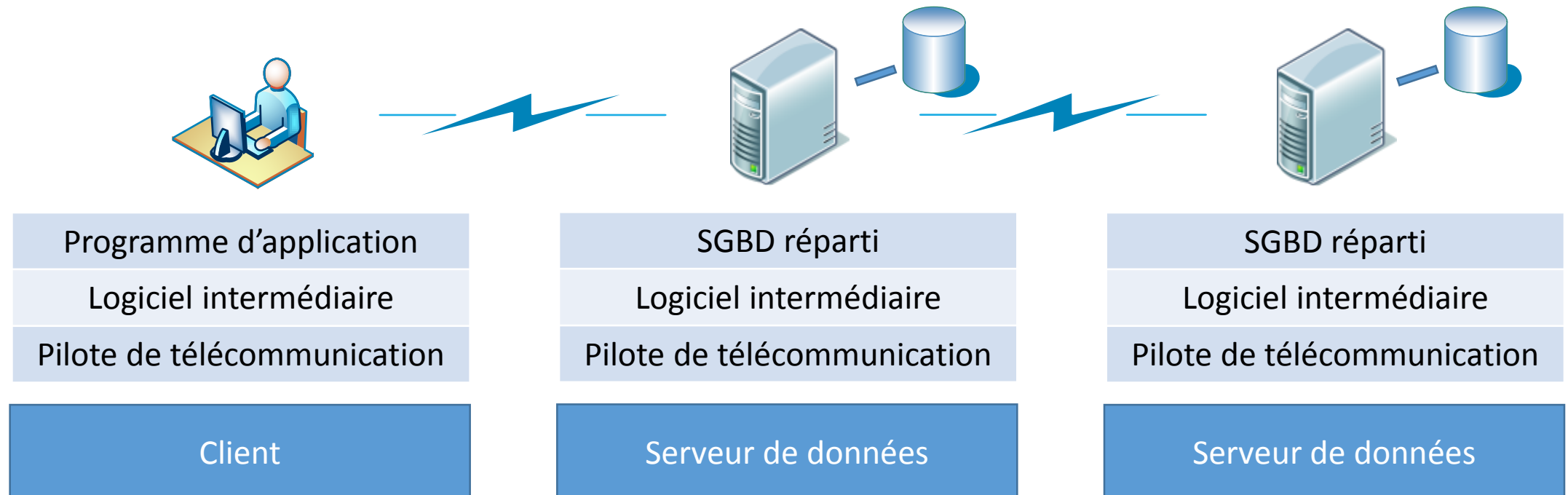
L'architecture d'un SGBD

- Une architecture 3 tiers



L'architecture d'un SGBD

- Une architecture distribuée



Le modèle relationnel

Le domaine

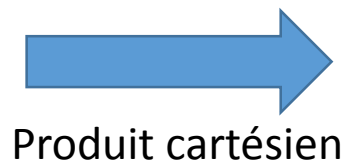
Un domaine représente un ensemble fini ou infini de valeurs.
Ces valeurs peuvent être de tout types.

Domaine de l'alphabet	Domaine des booléens	Domaine des doigts de pied	Domaine des nombres	Domaine des prénoms
A	0	Hallux
B	1	Secundus	-99999	Marie
C		Tertius	..	Jean
D		Quartus	0	Nathalie
...		Quintus	...	Philippe
Y			99999	Isabelle
Z		

Le produit cartésien

Pour tout ensemble A et tout ensemble B, il existe un ensemble, unique, dont les seuls éléments sont tous les couples dont la première composante appartient à A et la seconde à B.

Domaine des couleurs	Domaine des valeurs	
Pique	A	R
Cœur	D	V
Carreau	10	9
Trèfle	8	7
	6	5
	4	3
	2	



```
{
  (A, pique) ... (2, pique) ,
  (A, cœur) ... (2, cœur) ,
  (A, carreau) ... (2, carreau) ,
  (A, trèfle) ... (2, trèfle)
}
```

Un attribut

Un attribut est une propriété permettant de caractériser une entité. Les attributs possèdent un domaine de valeurs.



La valeur est un attribut de l'entité carte

La relation (ou entités)

Une relation est un sous-ensemble du produit cartésien d'une liste de domaines caractérisé par un nom.

Produit cartésien couleurs x valeurs

As de pique		2 de pique		3 de pique	
A	Pique	2	Pique	3	Pique
As de coeur		2 de coeur		3 de coeur	
A	Coeur	2	Coeur	3	Coeur
As de trèfle		2 de trèfle		3 de trèfle	
A	Trèfle	2	Trèfle	3	Trèfle
As de carreau		2 de carreau		3 de carreau	
A	Carreau	2	Carreau	3	Carreau

...

R de pique	
R	Pique
R	Coeur
R	Trèfle
R	Carreau

Relation appelée « R de pique »
dont les attributs sont :

- R
- Pique

La clé

Une clé de relation est un sous-ensemble qui permet de caractériser tout enregistrement d'une relation.

Id commande	Id objet	quantité
1	x3452	4
2	x3452	13
...	...	
4358	LXRH13402	1

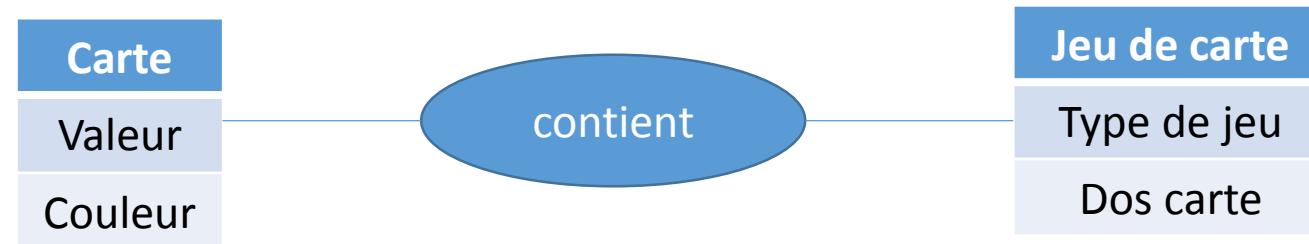
Une commande ne peut être définie que par un identifiant unique créé pour l'occasion.

Couleur	Valeur
A	Pique
2	Pique
...	...
R	Carreau

Une carte doit obligatoirement être définie par sa couleur et valeur.

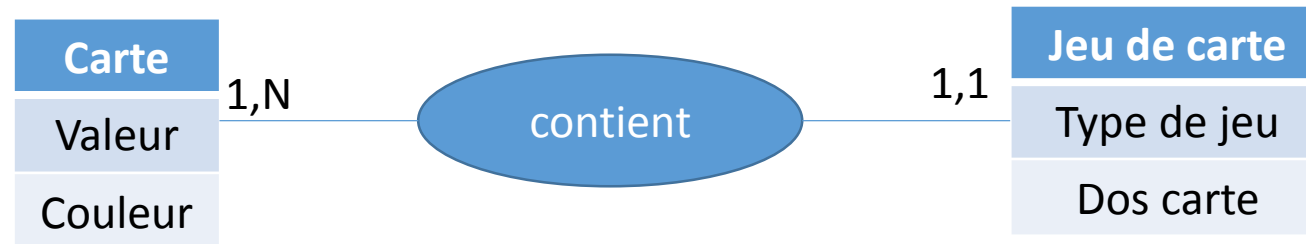
L'association

- Une association est une relation entre plusieurs entités. Elle permet de donner un sens métier aux entités dans le schéma général.

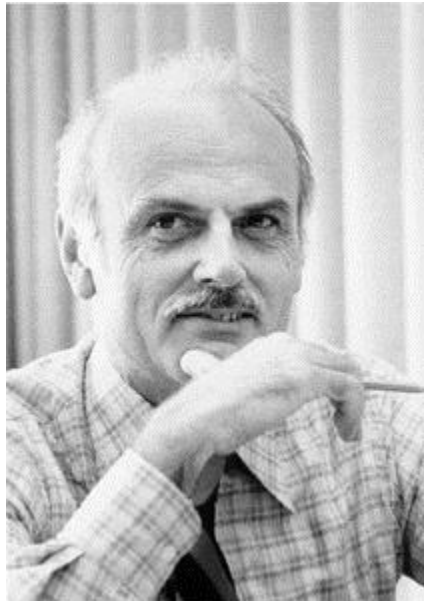


La cardinalité

- Une cardinalité relie une association et une entité en précisant le nombre de fois minimal et maximal d'interventions de l'entité dans l'association.



L'algèbre relationnel



Edgar Frank Codd

Problématique

Comment créer une information à partir de données?

Solution

Permettre à l'utilisateur de manipuler les données via un langage mathématique simplifié.

Exemple

Problématique

Je souhaite connaître le nom et prénom des individus dont la taille est supérieur à la moyenne de son genre.

J'ai pour cela un échantillon de 1000 personnes.

Pour chaque personne, je connais :

- Son prénom;
- Son nom;
- Sa taille;
- Son genre;

Comment faire?

Les opérateurs relationnels

L'union

Notation

$R \cup S$

Sens

$R \cup S = T$ tel que pour tout $t \in T$, il existe une relation identique dans R ou S.

Exemple

Je souhaite rassembler deux sets de données :

Personne 1

Prénom	Nom	Taille	Sexe
Pierre	Or	1.80	m
Paul	Ni	1.70	m

Personne 2

Prénom	Nom	Taille	Sexe
Jacqueline	Car	1.60	f

$Personne1 \cup Personne2$

Prénom	Nom	Taille	Sexe
Pierre	Or	1.80	m
Paul	Ni	1.70	m
Jacqueline	Car	1.60	f

Les opérateurs relationnels

L'intersection

Notation

$R \cap S$

Sens

$R \cap S = T$ tel que pour tout $t \in T$, il existe une relation identique dans R et S.

Exemple

Je souhaite trouver la relation se trouvant dans les deux sets de données ci-dessous:

Personne 1

Prénom	Nom	Taille	Sexe
Pierre	Or	1.80	m
Paul	Ni	1.70	m

Personne 2

Prénom	Nom	Taille	Sexe
Pierre	Or	1.80	m
Paul	No	1,70	m



$Personne1 \cap Personne2$

Prénom	Nom	Taille	Sexe
Pierre	Or	1.80	m

Les opérateurs relationnels

La restriction

Notation

$\sigma_F(R)$

σ se prononce Sigma

Sens

$\sigma_F(R) = \{ r \in R : r \text{ satisfait la condition donnée par } F \}$

Exemple

Je souhaite connaître les individus de sexe masculin de l'échantillon suivant nommé « Personne » :

Personne

Prénom	Nom	Taille	Sexe
Pierre	Or	1.80	m
Paul	Ni	1.70	m
Jacqueline	Car	1.60	f



Prénom	Nom	Taille	Sexe
Pierre	Or	1.80	m
Paul	Ni	1.70	m

$\sigma_{\text{sexe}='m'}(\text{Personne})$

Les opérateurs relationnels

La projection

Notation

$\pi_A(R)$

π se prononce pi

Sens

$\pi_A(R)$ qui est le relation R où on ne considère que les attributs de A

Exemple

Je ne souhaite connaître que les noms et prénoms de l'échantillon « Personne » :

Personne

Prénom	Nom	Taille	Sexe
Pierre	Or	1.80	m
Paul	Ni	1.70	m
Jacqueline	Car	1.60	f



Prénom	Nom
Pierre	Or
Paul	Ni
Jacqueline	Car

$\pi_{\text{prénom,nom}}(\text{Personne})$

Les opérateurs relationnels

La jointure

Notation

 $R \bowtie_Q S$

Sens

L'union des relations R et S satisfaisant le critère Q

Exemple

Je souhaite connaître la taille des individus:

Patronyme			Caractéristique		
ID	Prénom	Nom	ID	Taille	Sexe
1	Pierre	Or	1	1.80	m
2	Paul	Ni	2	1.70	m
3	Jacqueline	Car	3	1.60	f



ID	Prénom	Nom	Taille	Sexe
1	Pierre	Or	1.80	m
2	Paul	Ni	1.70	m
3	Jacqueline	Car	1.60	f

$Patronyme \bowtie_{id = id} Caractéristique$

Les opérateurs relationnels

Rebaptiser

Notation

$\rho_{a/b}(R)$

ρ se prononce rhô

Sens

$\rho_{a/b}(R)$ est la relation R avec b rebaptisé a

Exemple

Je souhaite renommer le champ ID :

Personne

ID	Prénom	Nom	Taille	Sexe
1	Pierre	Or	1.80	m
2	Paul	Ni	1.70	m
3	Jacqueline	Car	1.60	f



Identifiant	Prénom	Nom	Taille	Sexe
1	Pierre	Or	1.80	m
2	Paul	Ni	1.70	m
3	Jacqueline	Car	1.60	f

$\rho_{\text{identifiant/id}}(\text{Personne})$

Complétude

- Notre algèbre est donc complète
 - Nous pouvons formaliser toutes questions s'inscrivant dans une logique de premier ordre (sans fonction)

Exemple

Je souhaite connaître le prénom des hommes dont la taille est supérieure à 1m80

Personne		
ID	Prénom	Nom
1	Pierre	Or
2	Paul	Ni
3	Jacqueline	Car
4	Jean	Martin
5	Lucie	Bernard
6	Julie	Dubois
7	Nicolas	Petit

Caractéristique		
ID	Taille	Sexe
1	1.80	m
2	1.70	m
3	1.60	f
4	1.54	m
5	1.82	f
6	1.75	f
7	1.95	m

Projection (Prénom,
Restriction (sexe='m' et taille > 1.80,
Jointure (Personne, Caractéristique, ID)))

Prénom
Pierre
Nicolas