II. Conception de BDR

2024

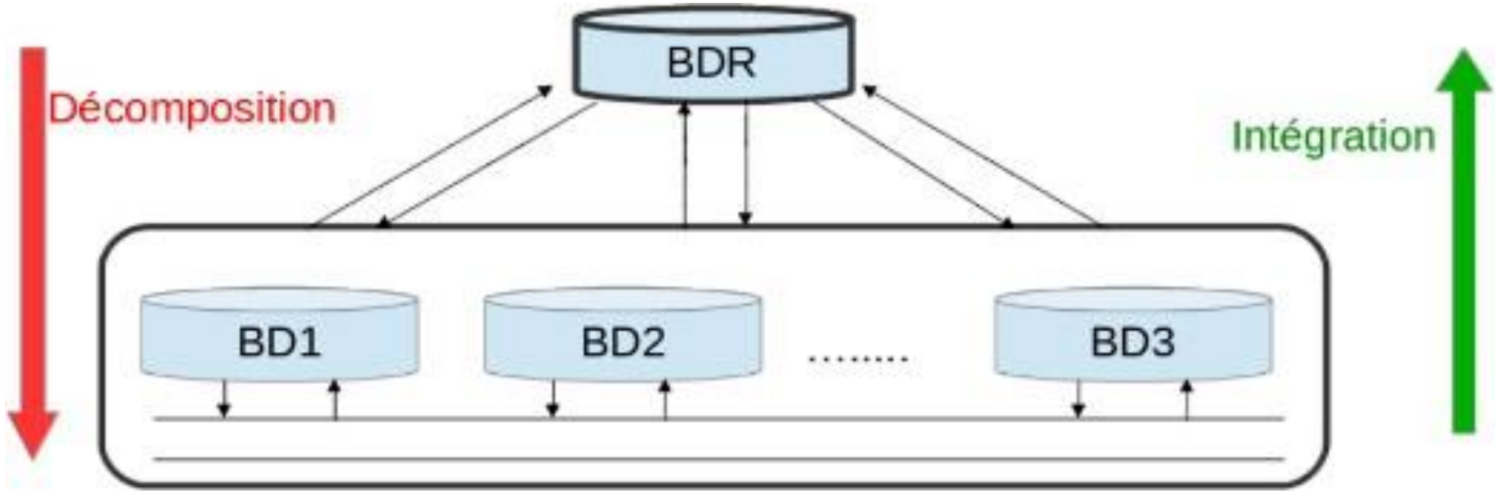
Introduction

La définition du schéma de répartition est la partie la plus délicate de la phase de conception d'une BDR, car **il n'existe pas de méthode** miracle pour trouver la **solution optimale**.

L'administrateur doit donc prendre des décisions dont l'objectif est de minimiser le nombre de transferts entre sites, les temps de transfert, le volume de données transférées, les temps moyens de traitement des requêtes, et le nombre de copies de fragments, ... etc.

Introduction

Pour faire la conception d’une base de données repartie, il existe deux approches : l’**approche ascendante** (button-up) et l’**approche descendante** (top-down).



I.1 Approche Ascendante pour la conception de BDR

L'approche se base sur l'existence de la répartition, mais il faut réussir à intégrer les différentes BDs existantes en une seule BD globale. En d'autres termes, les schémas conceptuels locaux existent et il faut réussir à les unifier dans un schéma conceptuel global.

I.2 Approche descendante pour la conception de BDR

On doit d'abord spécifier la liste des besoins pour la conception de la BD. Cette liste peut être classé en deux catégories:

1 Établissement des vues (interfaces, etc.).
2 Établissement des modèles conceptuels (E/A, en Classes).

Le schéma conceptuel global (GCS) résultant des étapes précédentes est utilisé pour concevoir la distribution du système.

L'objectif sera alors l'établissement du schéma conceptuel local (LCS) en distribuant les entités sur les sites du système en utilisant la fragmentation de relations.

I.3. Architecture d’une BDR

Schema Externe

→ Définitions de la représentation externe (vis-à-vis de l’utilisateur)

Schema Conceptuel Global

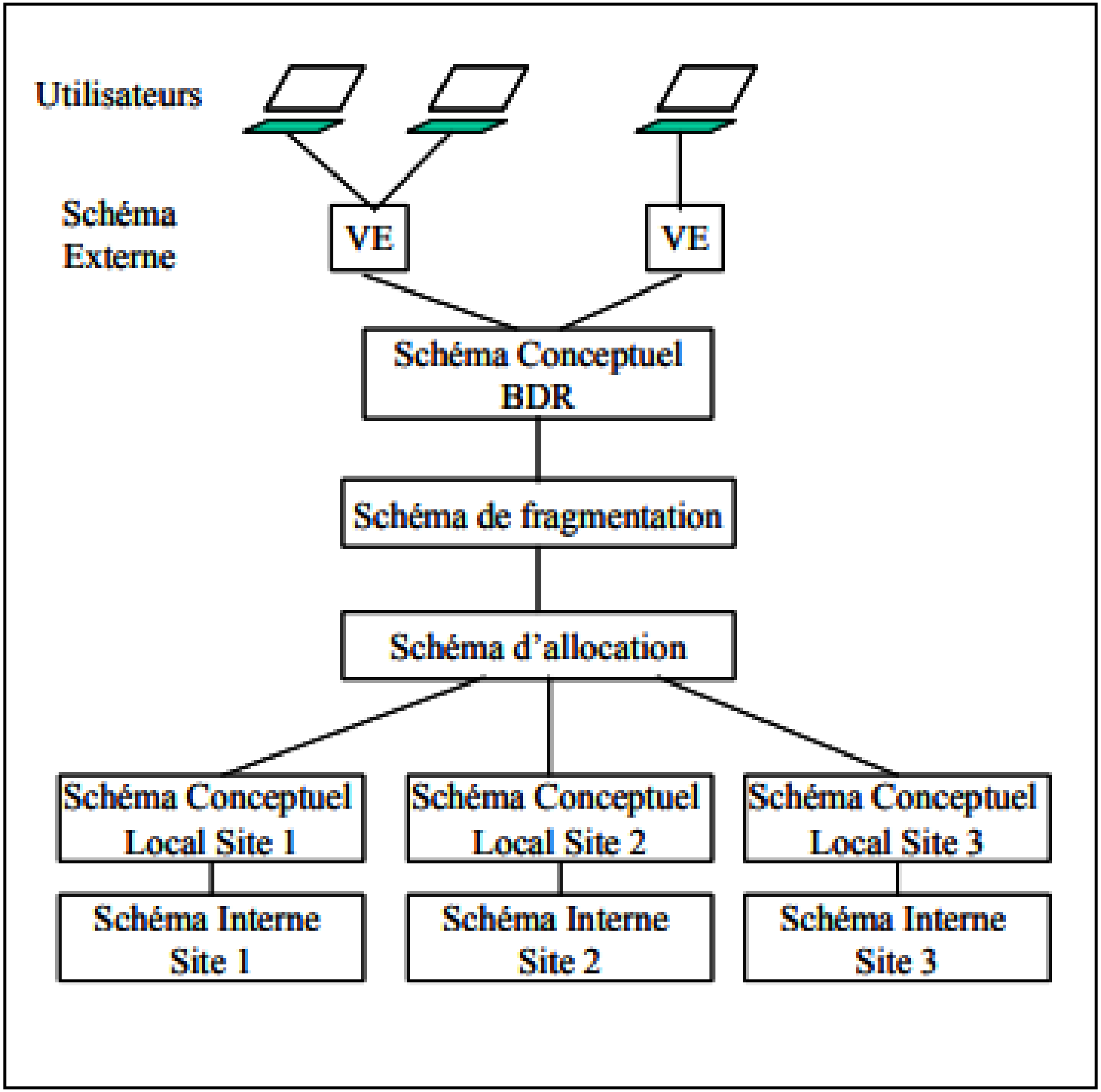
→ Schéma global des données dans la BDR.

Schema Conceptuel Local

→ Schéma local des données dans chaque BD locale.

LIS: Schema Interne Local

→ Organisation interne (physique) des données.



Architecture d'une BDR.

II. Conception Descendante d’une BDR

L'activité de conception de la distribution se résume à la **fragmentation** et à l'**allocation**

II.1 Fragmentation

La fragmentation est le processus de décomposition d'une base de données logique (telle qu'elle est vue par les utilisateurs) en un ensemble de "sous" bases de données (physiques). Cette décomposition doit évidemment être **sans perte d'information**

Une décomposition d'une relation globale *R* en petites relations appelés fragments: F_1, F_2, ..., F_n

a. Critères d'une bonne fragmentation

- **Complétude:**

  Une décomposition d'une relation *R* en fragments: $F_1, F_2, \ldots, F_n$ est complète si et seulement si chaque item de donnée dans *F* peut être aussi trouvé dans un des fragments F_i.

$$R \Rightarrow \mathbf{F} = \{F_1, F_2, \ldots\}$$

$$\forall t \in R,\ \exists\, F_i \in \mathbf{F} \quad \text{Tel que} \quad t \in F_i$$

- **Reconstruction:**

  Si une relation *R* est décomposée en fragments: $F_1, F_2, \ldots, F_n$, alors il doit exister un opérateur ∇ tel que: $R = \nabla_{1 \le i \le n} F_i$.

$$\exists\, \nabla \quad \text{Tel que} \quad R = \overset{i}{\nabla} F_i$$

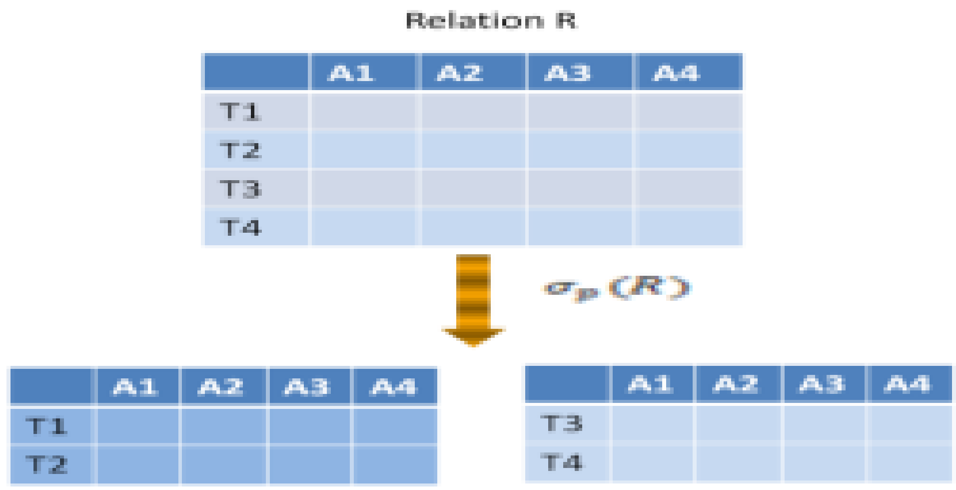
a. Critères d'une bonne fragmentation

- **Disjonction:**

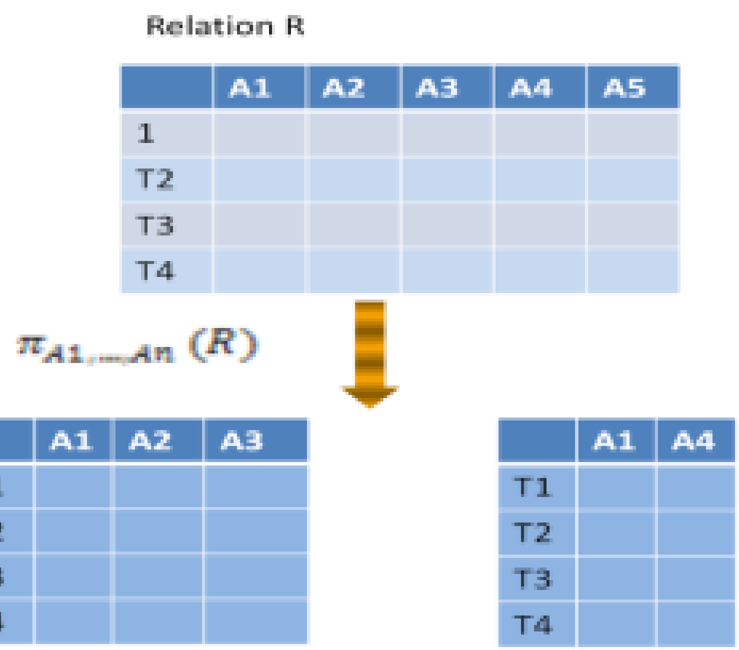
  Si une relation *R* est décomposée en fragments: $F_1, F_2, ..., F_n$, et que l'item $d_i \in F_j$, alors d_i ne doit pas être dans n'importe quel autre fragment F_k (k ≠ j).

$$F_i \cap F_j = \emptyset, \ \forall\, i,j \quad \text{Tel que} \quad i \neq j$$

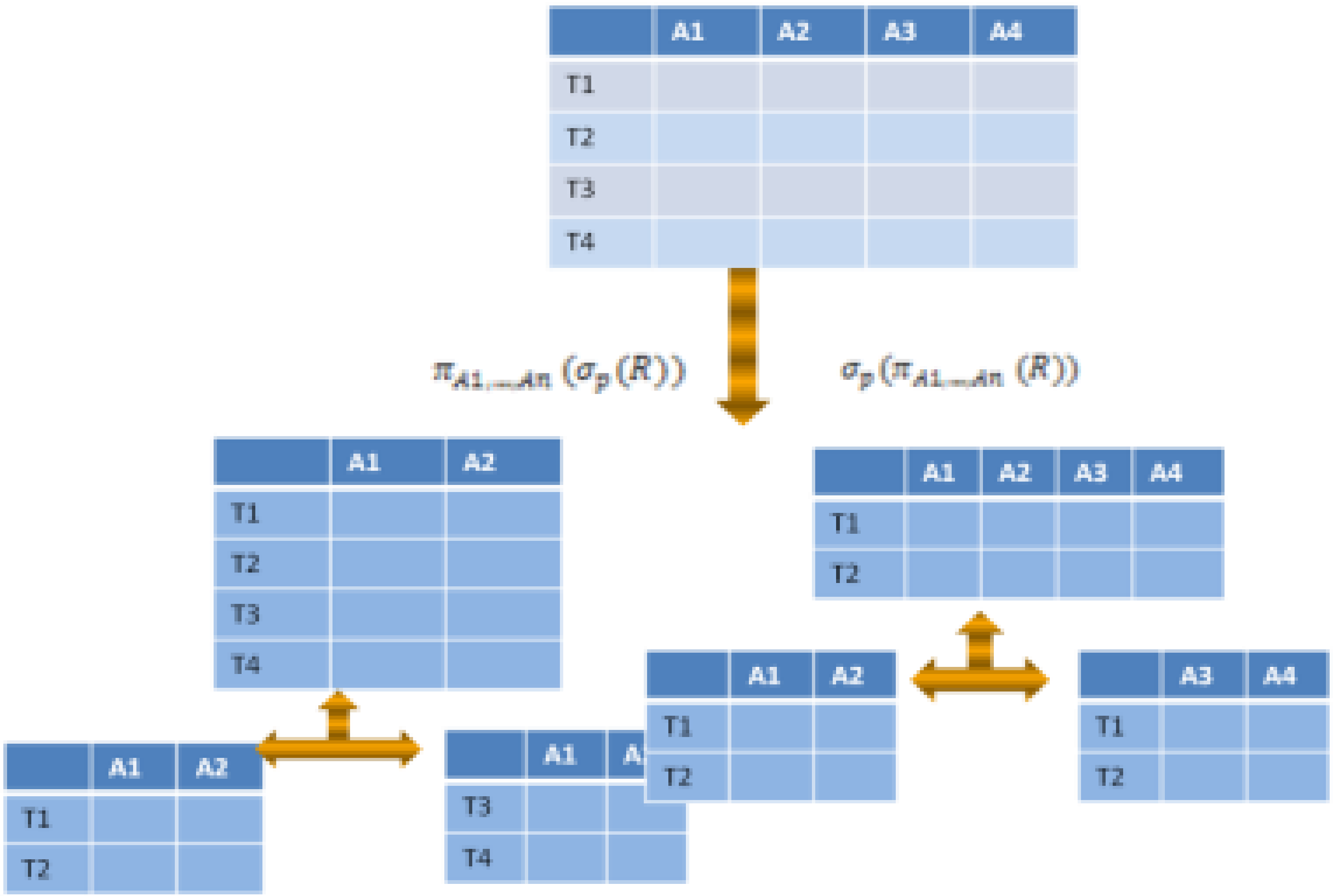
b. Alternatives de fragmentation:



Fragmentation horizontale d'une relation



Fragmentation verticale d'une relation



Fragmentation mixte d’une relation

c.Caractéristiques de la fragmentation

- **Fragmentation de relation (sous-relations)**
  - Permet l'exécution concurrente de plusieurs transactions.
  - Permet l'exécution parallèle de requêtes sur des fragments.
  - Les vues ne pouvant être définies sur un seul fragment requerront un traitement supplémentaire (jointures).

- Difficulté de vérifier l'intégrité sémantique.

III. Fragmentation horizontale (FH)

III. Fragmentation horizontale (FH)

Cette fragmentation partitionne les tuples d’une relation.

Schéma(R) = Schéma(F1) = Schéma(F2) = Schéma(F3) ..... Schéma(Fn)

Il existe deux types de la FH:

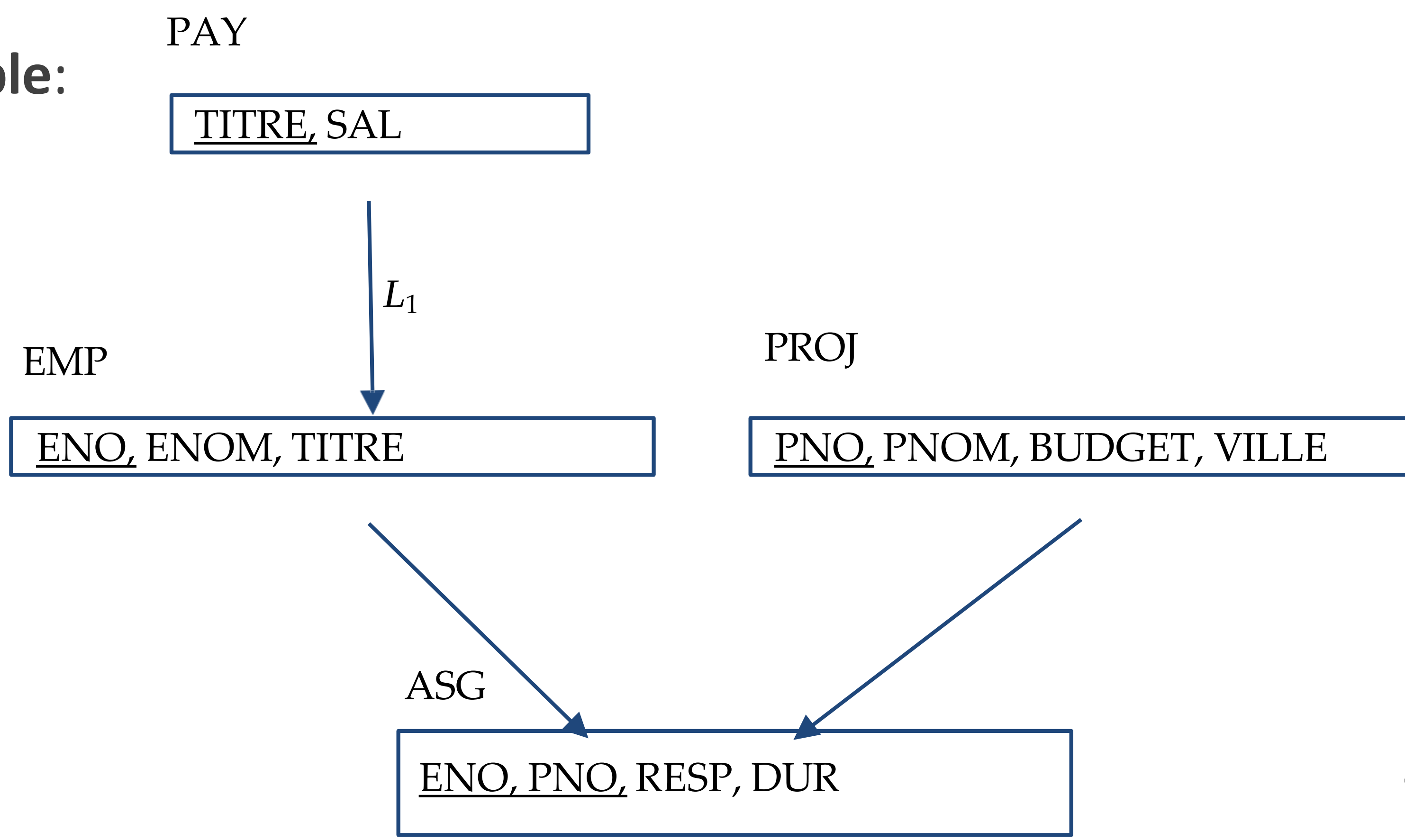
- **FH primaire**:
  En utilisant des prédicats définis sur la relation fragmentée.

- **FH dérivée:**
  En utilisant des prédicats définis sur une autre relation.

FH Primaire

Concerne le schéma conceptuel global (SCG). Il est important de savoir comment les relations sont interconnectées (avec les jointures).

Exemple:



Principe de FH Primaire

Pour chaque relation simple dans le schéma conceptuel global :

- Identifier les applications les plus importantes accédant à la relation
- Définir les **prédicats simples** selon les critères de sélections des requêtes définies sur la relation
- Générer les **prédicats minterms** comme conjonction des prédicats simples (*Ne pas oublier les formes négatives des prédicats simples*)
- Identifier les prédicats contradictoires et redondants, puis les éliminer
- Définir les **fragments** à partir d'opérations de sélections selon les prédicats minterms

FH Primaire

- **Prédicats simples**

Soit une relation $R[A_1, A_2, \ldots, A_n]$.

Un prédicat simple p_j sur R est define comme suit:

$$p_j : A_i \; \theta \; Valeur$$

Où $\theta \in \{=,<,\leq,>,\geq,\neq\}$, Valeur $\in D_i$ où D_i est le domaine de A_i.

Pour une relation R, on définit l'ensemble suivant de prédicats simples $P_r = \{p_1, p_2, \ldots, p_m\}$.

Exemple :

PNOM = "Maintenance"

BUDGET ≤ 200000

PROJ

PNO	PNOM	BUDGET	VILLE
P1	Conception	150000	JIJEL
P2	Dévelop BD	135000	BEJAIA
P3	Implémentation	250000	BEJAIA
P4	Maintenance	310000	ALGER

FH Primaire

➢ Prédicats mintermes

Soit une relation $R[A_1, A_2, \ldots, A_n]$, et un ensemble de prédicats simples $P_r = \{p_1, p_2, \ldots, p_m\}$.

On définit l'ensemble des mintermes $M = \{m_1, m_2, \ldots, m_z\}$ par:

$$M = \{ m_i \mid m_i = \wedge_{pj \in Pr} \; p_j^* \}, 1 \leq j \leq m, 1 \leq i \leq z$$

Où : $p_j^* = p_j$ ou $p_j^* = \neg(p_j)$,

$\wedge$ est l'opérateur ET logique.

FH Primaire

Exemple:

Prédicats Simples:

p1: PNOM="Maintenance"

p2: BUDGET≤200000

PROJ

PNO	PNOM	BUDGET	VILLE
P1	Conception	150000	JIJEL
P2	Dévelop BD	135000	BEJAIA
P3	Implémentation	250000	BEJAIA
P4	Maintenance	310000	ALGER

Prédicats Mintermes:

m1: (PNOM="Maintenance“) ∧ (BUDGET≤200000)

m2: ¬ (PNOM="Maintenance") ∧ (BUDGET≤200000)

m3: PNOM= ("Maintenance“) ∧ ¬ (BUDGET≤200000)

m4: ¬(PNOM="Maintenance") ∧ ¬ (BUDGET≤200000)

➢ Définition des fragments

Dans une relation R, les fragments F_j sont donnés par:

$$F_j = \sigma_{c_j}(R), \qquad 1 \leq j \leq w$$

Où c_j est une formule de sélection qui est de préférence un prédicat minterme. Il s'en suit qu'un fragment F_j d'une relation R consiste en tous les tuples de R qui satisfont le prédicat m_j.

$\Downarrow$

Soit un ensemble de prédicats M. Il existe autant de fragments horizontaux d'une relation R qu'il existe de prédicats mintermes.

L'ensemble de fragments horizontaux est appelé fragments mintermes.

➢ Définition des fragments

Il est clair que la définition de fragments horizontaux dépend des prédicats mintermes.

→ De ce fait, la première étape dans tout processus de fragmentation est de déterminer un ensemble de prédicats simples qui formeront les prédicats mintermes.

Deux aspects importants des prédicats simples sont:

- La complétude;
- La minimalité.

➢ complétude

Un ensemble de prédicats P_r simples est dit complet si et seulement si il existe une probabilité égale d'accès de toutes les applications pour n'importe quel fragment défini par P_r.

Exemple: Soit la relation PROJ(PNO, PNOM, BUDGET, VILLE) et soit l'ensemble des prédicats:

Pr = {Ville="Béjaia", Ville="JIJEL", Ville="ALGER"}

Si la **seule application** qui accède à la relation PROJ le fait selon **la localisation**, alors l'ensemble Pr est complet

- **complétude**

Soient maintenant deux applications définies comme suit:

1 Trouver les budgets de projets dans chaque ville.

2 Trouver les projets avec des budgets ≤ 200 000.

L'ensemble **Pr =** {ville="Alger", ville="Jijel", ville="Bejaia"} n'est pas complet, car certains tuples des fragments ont plus de probabilité d'être accédés par la 2e application.

→ Pr = {ville="Alger", ville="Jijel", ville="Bejaia" , BUDGET ≤ 200000, BUDGET > 200000} est complet car il offre une probabilité égale pour tous les tuples d'être accédés.

➢ complétude

L'objectif de la complétude d'un l'ensemble de prédicats est que les fragments obtenus par un tel ensemble sont logiquement uniformes.

$\Longrightarrow$ L'accès à ces derniers par les applications est statistiquement homogène.

La fragmentation résultant d'un ensemble de prédicats complet garantit un certain équilibre de la charge des fragments de la BDD sur les sites.

Il est donc primordial de faire la FH par un ensemble de prédicats complet.

➢ minimalité

La deuxième propriété désirable de l’ensemble des prédicats est la minimalité.

Si un prédicat simple influe sur la façon dont la fragmentation est effectuée, (c.-à-d. provoque une division d’un fragment f en fragments f_i et f_j), il devrait y avoir au moins une application qui accède à f_i et f_j différemment.

En d'autres termes, un simple prédicat devrait être pertinent pour déterminer une fragmentation.

Si tous les prédicats d'un ensemble P_r sont pertinents, on dit que P_r est minimal.

➢ minimalité

Exemple:

L'ensemble de prédicats P_r = {ville="Alger", ville="Jijel", ville="Bejaia" , BUDGET ≤ 200000, BUDGET > 200000} est minimal en plus d'être complet.

Cependant, si on rajoute le prédicat PNOM = "Conception" à P_r, l'ensemble devient non minimal, car le nouveau prédicat est non pertinent (c.-à-d., il n'existe pas d'application qui accèderait au fragments résultants de manière différente)

➢ Elimination des mintermes contradictoires

On va devoir réduire le nombre de mintermes pouvant être considérés dans la fragmentation. Cette élimination peut être faite en identifiant des mintermes pouvant être contradictoires à un ensemble d'implications **I**:

Exemple:

$P_r' = \{p_1, p_2\}$ où: P_1: att = val_1
p_2: att = val_2

L'ensemble **I** dans ce cas peut être:

i_1: (att = val_1) ⇒ ¬(att = val_2)
i_2: ¬(att = val_1) ⇒ (att = val_2)

➢ Elimination des mintermes contradictoires

Exemple (suite):

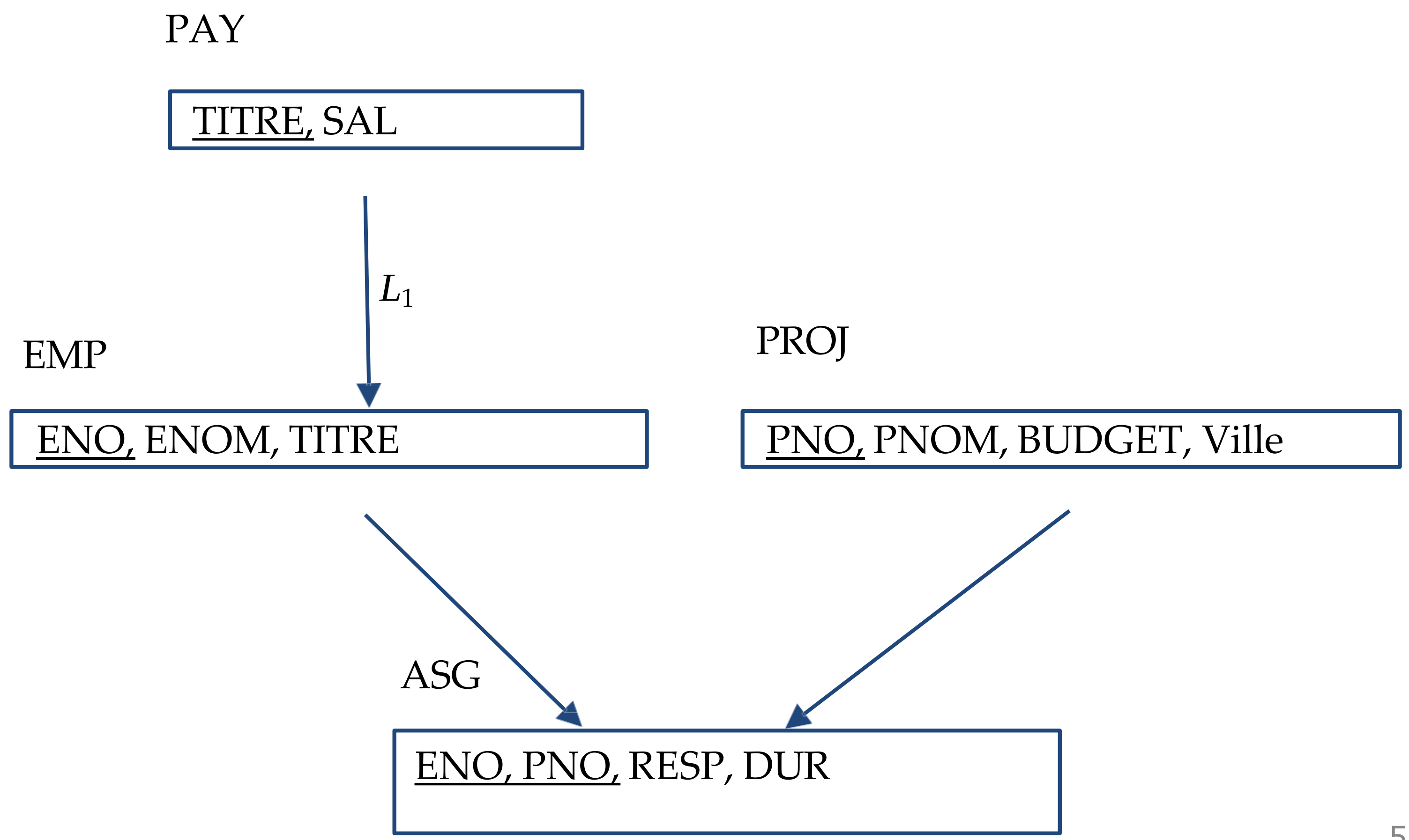
Les 4 mintermes suivants peuvent être formés à partir de l'ensemble des prédicats $P_r' = \{p_1, p_2\}$:

m_1: (att = val_1) $\wedge$ (att = val_2)
m_2: (att = val_1) $\wedge$ $\neg$(att = val_2)
m_3: $\neg$(att = val_1) $\wedge$ (att = val_2)
m_4: $\neg$(att = val_1) $\wedge$ $\neg$(att = val_2)

Dans ce cas, les mintermes $\mathbf{m_1}$ et $\mathbf{m_4}$ sont contradictoires par rapport à l'ensemble d'implications précédent $\mathbf{I}$.

a. FH primaire: Exemples

Soit la BD



FH primaire: Exemple 1

On peut appliquer une FH primaire sur les relations PAY et PROJ. On suppose qu'il existe une application qui accède à la relation PAY en regardant le salaire et l'incrémentant selon son montant.

On suppose que les enregistrements des employés sont gardés dans 2 sites. Un des sites garde les employés avec des salaires ≤ 30 000 , l'autre site garde les employés avec des salaires > 30 000.

Les prédicats qui peuvent servir à partitionner PAY sont alors: p_1:

SAL ≤ 30 000

p_2: SAL > 30 000.

FH primaire: Exemple 1

En ayant l'ensemble initial de prédicats $P_r = \{p_1, p_2\}$, on obtient $P_r' = \{p_1\}$ qui est complet et minimal (en effet, p_2 ne pourrait partitionner f_1 résultant de p_1.

On pourrait alors former l'ensemble M des mintermes suivant:

m_1: SAL $\leq$ 30 000.

m_2: $\neg$ (SAL $\leq$ 30 000).

Les fragments obtenus par M sont les suivants:

PAY_1

TITRE	SAL
Ing.Mech. .	27000
Programmeur	24000

PAY_2

TITRE	SAL
Ing. Elect.	40000
Anal. Syst.	34000

FH primaire: Exemple 2

Considérons maintenant la relation PROJ. On suppose qu'il existe 2 applications l'accédant.

La première application est originaire de 3 sites et a le rôle de trouver les noms et les budgets de projets en ayant leur localisation. En SQL, une telle requête s'écrit:

```
SELECT    PNAME, BUDGET
FROM      PROJ
WHERE     Ville = 'valeur'
```

Pour cette application, les prédicats à utiliser sont:

p_1: Ville="Béjaia"
p_2: Ville="JIJEL"
p_3: Ville="ALGER"

FH primaire: Exemple 2

La 2e application s'occupe de la gestion des projets et est originaire de 2 sites. Les projets ayant un budget ≤ 200 000 sont gérés sur un site et les autres autre site. Les prédicats qui doivent être utilisés pour cette application sont comme suit:

p_4: BUDGET ≤ 200 000.

p_5: BUDGET > 200 000.

Alors, l'ensemble P_r'={p_1, p_2, p_4} est complet est minimal.

En utilisant P_r', l'ensemble des mintermes suivants peut être défini:

m_1: (Ville = 'Jijel') ∧ (BUDGET ≤ 200 000)

m_2: (Ville = ' jijel') ∧ (BUDGET > 200 000)

m_3: (Ville = 'Bejaia') ∧ (BUDGET ≤ 200 000)

m_4: (Ville = 'Béjaia') ∧ (BUDGET > 200 000)

m_5: (Ville = 'Alger') ∧ (BUDGET ≤ 200 000)

m_6: (Ville = 'Alger') ∧ (BUDGET > 200 000)

a. FH primaire: Exemple

PROJ

PNO	PNOM	BUDGET	VILLE
P1	Conception	150000	JIJEL
P2	Dévelop BD	135000	BEJAIA
P3	Implémentation	250000	BEJAIA
P4	Maintenance	310000	ALGER

$PROJ_1$

PNO	PNOM	BUDGET	Ville
P1	Conception	150000	JIJEL

$PROJ_2$

PNO	PNOM	BUDGET	Ville
P2	Develop BD.	135000	BEJAIA

$PROJ_4$

PNO	PNOM	BUDGET	Ville
P3	Implémentation	250000	BEJAIA

$PROJ_6$

PNO	PNOM	BUDGET	Ville
P4	Maintenance	310000	ALGER

FHP: Vérification des critères

- **Complétude**

Pour la FH primaire, si l'ensemble des prédicats P_r' est complet et minimal, on garantit que le critère de **complétude** de la fragmentation est satisfait.

- **Reconstruction**

La **reconstruction** d'une relation globale à partir de ses fragments est effectuée avec l'opérateur d'union dans la FH primaire
Pour une relation R avec une fragmentation FR = $\{R_1, R_2, \ldots, R_w\}$, on aura :

- **Disjonction**

- La **disjonction** est garantie dans la FH primaire lorsque les prédicats mintermes sont mutuellement exclusifs.

FH dérivée

Une FH dérivée est définie sur une relation membre d'un lien selon une opération de sélection spécifiée sur sa relation principale.

Le lien entre une relation principale et une relation membre est défini par opération d'équijointure.

On voudrait partitionner une relation membre selon la fragmentation de la relation principale, mais en utilisant les attributs de la relation membre.

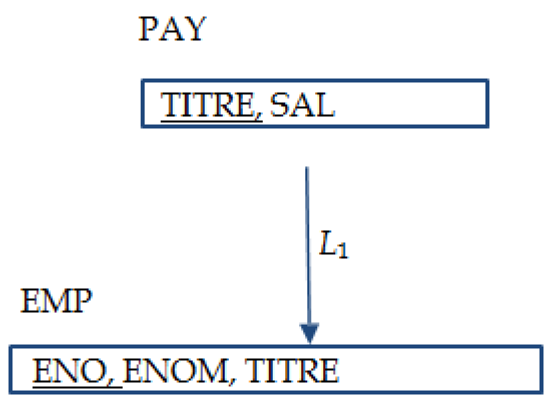
Soit un lien **L** où PRINCIPALE(**L**) = S et MEMBRE(**L**) = R.

FH dérivée

Les fragments horizontaux dérivés de R sont définis par:

$R_i = R \ltimes S_i, \;\; 1 \leq i \leq w$

Où w est le nombre maximal de fragments qui seront définis sur R, et $S_i = \sigma_{c_i}(S)$, où c_i est la formule selon laquelle le fragment horizontal primaire S_i est défini.



Exemple: Soit le lien $\mathbf{L_1}$ de la BD entreprise. On a:

PRINCIPALE($\mathbf{L_1}$) = PAY, MEMBRE($\mathbf{L_1}$) = EMP.

On peut regrouper les employés en 2 groupes selon le salaire:

FH dérivée

Ceux qui ont des salaires ≤ 30 000 et ceux ayant des salaires > 30 000.
Les deux fragments de **EMP** sont définis comme suit:

$EMP_1 = EMP \ltimes PAY_1$
$EMP_2 = EMP \ltimes PAY_2$

Où: $PAY_1 = \sigma_{(SAL \leq 30\,000)}(PAY)$; $PAY_2 = \sigma_{(SAL > 30\,000)}(PAY)$

EMP_1

ENO	ENOM	TITRE
E3	A. Mohammed	Ing.Mach. .
E4	J. Yassine	Programmeur
E7	R. Salim	Ing. Mach

EMP_2

ENO	ENOM	TITRE
E1	J. Anis	Ing. Elect.
E2	M. Adem	Anal. Syst.
E5	B. Fares	Anal. Syst.
E6	L. Sami	Ing. Elect.
E8	J. Yasser	Anal. Syst.

FH: Vérification des critères

➢ **Complétude**

soit PRINCIPALE(**L**) = S et MEMBRE(**L**) = R d'un lien **L**, et supposons que R et S sont fragmentées somme suit:

$F_S = \{S_1, S_2, \ldots, S_w\}$ et $F_R = \{R_1, R_2, \ldots, R_w\}$, respectivement.

Soit A l'attribut de jointure entre R et S. Alors, pour chaque tuple *t* de R, il doit exister un tuple *t'* de S_i, tel que *t*[A] = *t'*[A].

FH: Vérification des critères

Reconstruction:

La reconstruction d'une relation globale à partir de ses fragments est effectuée avec l'opérateur d'union dans FH primaire et dérivée.

Pour une relation R avec une fragmentation $F_R = \{R_1, R_2, \ldots, R_w\}$, on aura ce qui suit:

$$R = \bigcup R_i \ , \forall R_i \in F_R$$

FH: Vérification des critères

- **Disjonction**

Dans la FH dérivée, la disjonction est garantie si le graphe de jointure est simple:

→ En général, on ne voudrait pas qu'un tuple d'une relation membre soit joint avec deux tuples ou plus dans la relation principale, lorsque ces deux derniers sont dans 2 fragments différents

FH: Vérification des critères

Disjonction (suite):

Exemple:

Dans la fragmentation de la relation PAY, les prédicats mintermes $M=\{m_1, m_2\}$ sont m_1: SAL $\leq$ 30 000 et m_2: SAL > 30 000. Vu que m_1 et m_2 sont mutuellement exclusifs, la fragmentation de PAY est disjointe.

Pour la fragmentation de EMP, on aura besoin que:

1 Chaque Employé possède un seul titre.

2 Chaque titre est relié à une seule valeur de salaire.