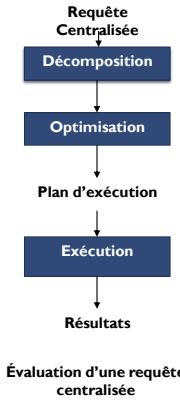
Évaluation des Requêtes Distribuées

TRAITEMENT DES REQUÊTES CENTRALISÉE



Évaluation d'une requête centralisée

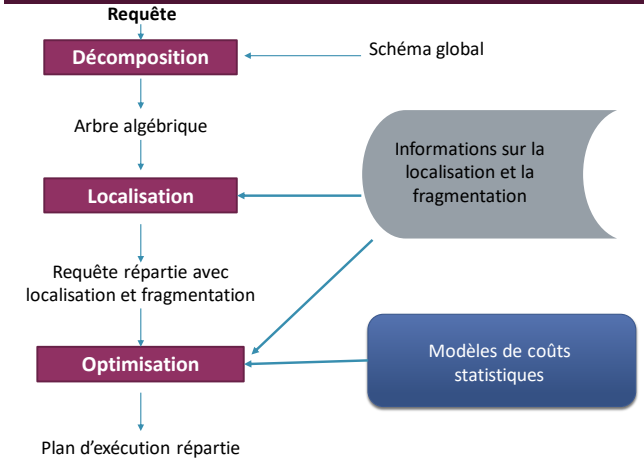
Décomposition *:* La requête SQL est transformée en arbre algébrique.

Optimisation : Ordonnancement optimal des opérations relationnelles, les algorithmes d'accès aux données en utilisant :

- **Les propriétés des opérateurs algébriques** (Associativité, Commutativité, Distributivité)
- **Des heuristiques** : en évaluant en premier les opérations les moins coûteuses le plutôt possible

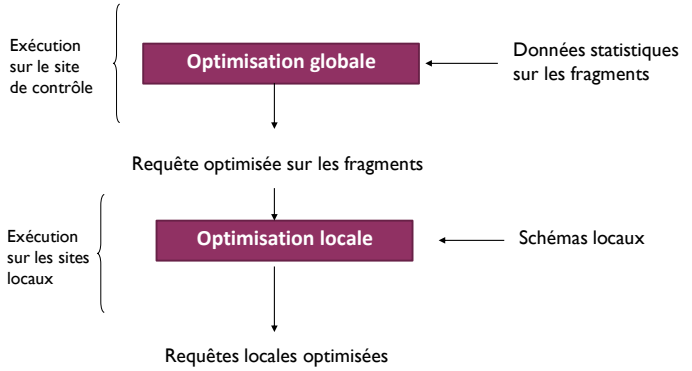
Exécution : Exécution de la requête optimisée afin d'élaborer son résultat.

TRAITEMENT DES REQUÊTES RÉPARTIES



3

OPTIMISATION DE REQUÊTES DISTRIBUÉES



4

OPTIMISATION DE REQUÊTES DISTRIBUÉES

1. **Décomposition de requête**: cette couche prend la requête exprimée en termes de relations globales et effectue une première optimisation partielle à l'aide d'heuristiques sur l'ordonnancement des opérateurs
2. **Localisation de données:** cette couche prend en compte la répartition des données sur le système. On remplace les feuilles de l'arbre d'algèbre relationnelle par leurs **algorithmes de reconstruction**
3. **Optimisation globale**: cette couche prend en compte les informations statistiques pour trouver un plan d'exécution proche de l'optimum, basé sur les fragments
4. **Optimisation locale**: cette couche s'exécute sur chacun des sites locaux impliqués dans la requête. Chaque SGBD local effectue ses propres optimisations à l'aide des heuristiques classiques d'optimisation

5

LA DÉCOMPOSITION

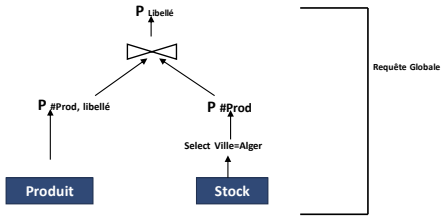
La requête globale est décomposée en sous-requêtes locales.

Relation :

Produit (#Prod, Libellé, Pu, …,)

Stock (#Prod,Ville, Qté)

La requête globale *Req* = Quels sont les produits stockés sur Alger ?



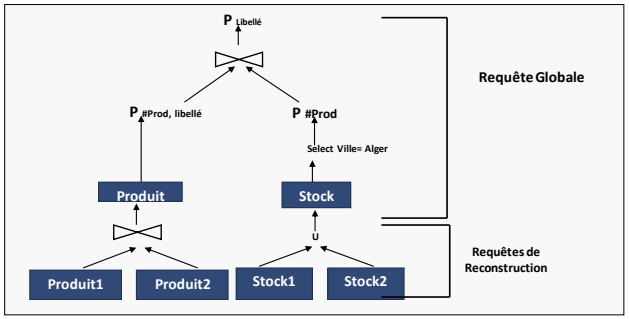
6

OPTIMISATION DE REQUÊTES: LA LOCALISATION

- **La localisation** d'une requête distribuée procède en deux étapes :
  - **Génération de requête canonique**
  - **Simplification (ou réduction)**
- *Génération de requête canonique*
  - la requête canonique est obtenue en remplaçant chaque relation de la requête distribuée globale par la requête **de reconstruction correspondante**
    - **Union** des fragments pour la fragmentation horizontale
    - **Jointure** des fragments pour la fragmentation Verticale

7

EXEMPLE



Site 1

Site 2

Produit1= **Proj** #prod,libellé (Produit)
Produit2 = **Proj** #prod,Pu (Produit)

Stock1 = **SEL** # Ville=Alger (Stock)
Stock2 = **SEL** # Ville <> Alger (Stock)

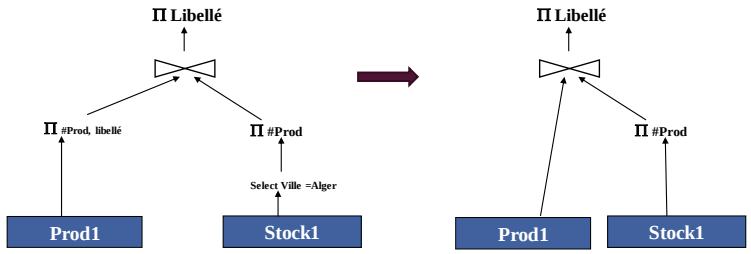
8

LA SIMPLIFICATION

- **Déterminer les opérations inutiles dans une requête canonique** : celles qui produisent un résultat vide, ou identique à l'opérande.
  - **Une opération de restriction** sur un fragment horizontal, dont le prédicat est en contradiction avec le prédicat de fragmentation, produit un résultat vide.
  - *Exemple: la restriction Ville= Alger au fragment Stock2 défini par Ville <> Alger*
- **Éliminer les conditions de sélections *inutiles :*** quand la condition est identique à celle de la fragmentation.
  - *Exemple: sélection à Ville=Alger sur le fragment Stock1 défini par Ville =Alger*
- **Une opération de projection sur un fragment vertical**, dont tous les attributs projetés- à l'exception, de l'attribut commun de reconstruction, n'appartiennent pas au fragment, donne un résultat vide.
  - *Exemple: projection sur numprod, libellé du fragment Produit2*

9

ARBRE FINAL



10

TRAITEMENT DES REQUÊTES RÉPARTIES

Rôle de l'optimisation : déterminer une stratégie d'exécution de la requête distribuée à moindre coût

- ***Le temps total d'exécution***
  - ***BD Centralisées*** *:Temps d'E/S + Temps de traitement*
  - ***BD Distribuées :*** *Temps d'E/S + Temps de traitement + Temps de communication (transferts de données, des résultats)*
  - *Prendre en considération les gains en cas de parallélisme (Max au lieu de la somme)*

11

OBJECTIFS ET FACTEURS DE L'OPTIMISATION

- Minimiser la fonction coût d'exécution : somme des temps d'exécution des sous requêtes sur chaque site
- **Tenir compte**
  - Du parallélisme
  - Des coûts de transferts
  - Des profils des fragments
    - Taille, nombre de n-uplets
    - Taille du domaine des attributs…
  - De la taille des résultats intermédiaires
  - De la topologie du réseau

12

TRAITEMENT DE REQUÊTES RÉPARTIES

L'exécution : le processus d'exécution de requêtes est aussi distribué, c'est-à-dire qu'il consistera en plusieurs processus d'exécution de requêtes locales et d'un processus d'exécution global.

- ***Le plan de l'exécution*** est l'ensemble des traitements locaux, ainsi que les opérations de communications des données intermédiaires, nécessaires pour les exécutions locales et la synchronisation globale.

13

TRAITEMENT DE REQUÊTES RÉPARTIES

- Les décisions à prendre dans le plan d'exécution sont les suivantes
  - Ordre des jointures
  - stratégie de jointure
  - Sélection des copies (site le plus proche, le moine engorgé)
  - Choix des sites d'exécution (fonction des coûts de communication)
  - Choix des algorithmes d'accès répartis
  - Choix du mode de transfert (tuple/paquets)

14

GESTION DU CATALOGUE

- Dans un système distribué, le catalogue du système contiendra non seulement les données d'un catalogue classique (relations de base, vues, index, …) mais également toutes les informations de contrôle nécessaires au système pour assure la fragmentation , la duplication..

Quelques possibilités

1. **Centralisé**
   - l'ensemble du catalogue est mémorisé une seule fois, sur un site central unique.
- ***Cette approche viole l'objectif « pas de contrôle centralisé »***
2. **Duplication totale**
   - l'ensemble du catalogue est entièrement mémorisé sur chaque site.

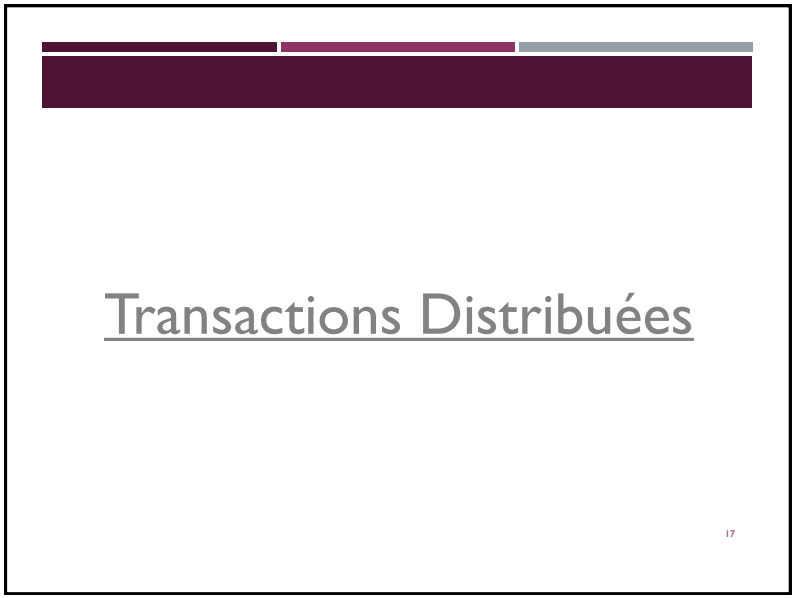
15

- ***Perte importante d'autonomie***

GESTION DU CATALOGUE

- **3. Partitionné**
  - Chaque site gère son propre catalogue concernant les objets de ce site. L'ensemble du catalogue est l'union de tous les catalogues locaux et disjoints.
- ***Les opérations qui ne sont pas locales sont très coûteuses (trouver un objet distant nécessite d'accéder en moyenne à la moitié des sites)***
- **Combinaison de 1 et 3**
  - Chaque site gère son propre catalogue local, de plus un site central gère une copie unifiée de tous les catalogues locaux.
- *Plus efficace que 3 mais viole l'objectif « pas de contrôle centralisé ».*

16


Transactions Distribuées
17