

Base de données distribuées



Babacar Diop

Dpt. d'Informatique

UFR des Sciences Appliquées et de Technologies

Université Gaston Berger de Saint-Louis

2018/2019

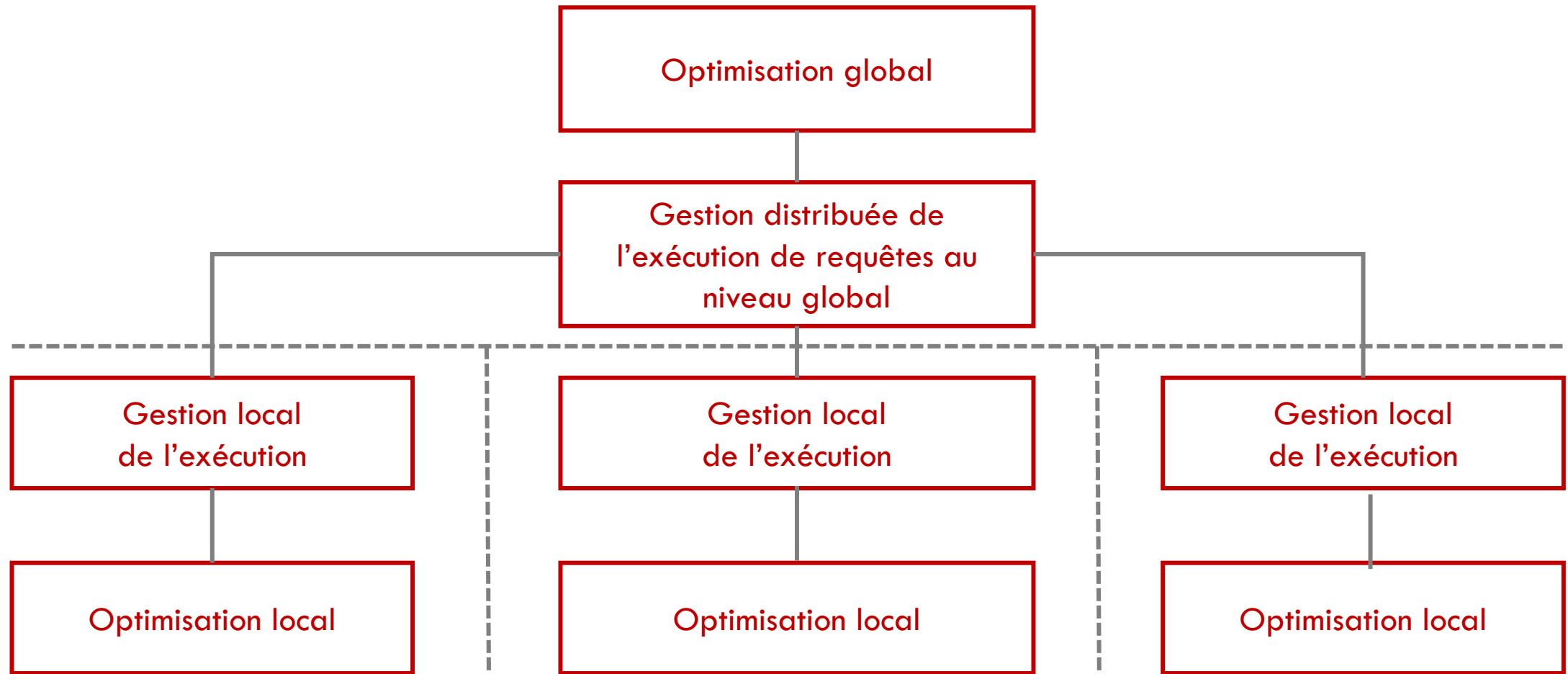
Gestion de requêtes dans une BD distribuée

Traitement de requêtes

dans une BD distribuée

- Dans un SGBDD, le traitement d'une requête comprend une optimisation à la fois au niveau global et au niveau local
- La requête entre dans le SGBDD sur le client ou sur le site de contrôle
- L'utilisateur est validé, la requête est vérifiée, traduite et optimisée au niveau global

Traitement de requêtes dans une BD distribuée



Traitement de requêtes

dans une BD distribuée

- Mapping de requêtes globales aux requêtes locales
- Le processus de mapping des requêtes globales aux requêtes locales peut être réalisé comme suit:
 - Les tables requises dans une requête globale ont des fragments répartis sur plusieurs sites
 - Les bases de données locales contiennent uniquement des informations sur les données locales
 - Le site de contrôle utilise le dictionnaire de données global pour rassembler des informations sur la distribution et reconstruire la vue globale à partir des fragments

Traitement de requêtes

dans une BD distribuée

- En l'absence de réplication, l'optimiseur global exécute des requêtes locales sur les sites où les fragments sont stockés
- S'il y a réplication, l'optimiseur global sélectionne le site en fonction des coûts de communication, de la charge de travail et de la vitesse du serveur
- L'optimiseur global génère un plan d'exécution réparti de manière à minimiser le transfert de données sur les sites
- Le plan indique l'emplacement des fragments, l'ordre dans lequel les étapes de la requête doivent être exécutées et les processus impliqués dans le transfert des résultats intermédiaires

Traitement de requêtes

dans une BD distribuée

- Par exemple, considérons que le schéma de projet suivant est fragmenté horizontalement selon les villes, New Delhi, Kolkata et Hyderabad.

PROJECT

PIId	City	Department	Status
------	------	------------	--------

Traitement de requêtes

dans une BD distribuée

Supposons qu'il existe une requête pour extraire les détails de tous les projets dont le statut est «En cours».

Requête globale	$\sigma_{status} = "ongoing"$ (PROJECT)
Requête sur new Dheli	$\sigma_{status} = "ongoing"$ (NewD-PROJECT)
Requête sur new Kolkata	$\sigma_{status} = "ongoing"$ (Kol-PROJECT)
Requête sur new Hyderabad	$\sigma_{status} = "ongoing"$ (Hyd-PROJECT)

Traitement de requêtes

dans une BD distribuée

- Pour obtenir le résultat global, nous devons regrouper les résultats des trois requêtes de la manière suivante:

$$\sigma_{status} = \text{"ongoing"}(NewD_PROJECT) \cup \sigma_{status} = \text{"ongoing"}(kol_PROJECT) \\ \cup \sigma_{status} = \text{"ongoing"}(Hyd_PROJECT)$$

Optimisation de requêtes

dans une BD distribuée

- L'optimisation de requête distribuée nécessite l'évaluation d'un grand nombre d'arbres de requête produisant chacun les résultats requis d'une requête
- Ceci est principalement dû à la présence d'une grande quantité de données répliquées et fragmentées
- Par conséquent, l'objectif est de trouver une solution optimale au lieu de la meilleure

Optimisation de requêtes

dans une BD distribuée

- Les principaux problèmes liés à l'optimisation des requêtes distribuées sont les suivants:
 - utilisation optimale des ressources dans le SD,
 - requête trading,
 - réduction de l'espace de solution de la requête.

Optimisation de requêtes

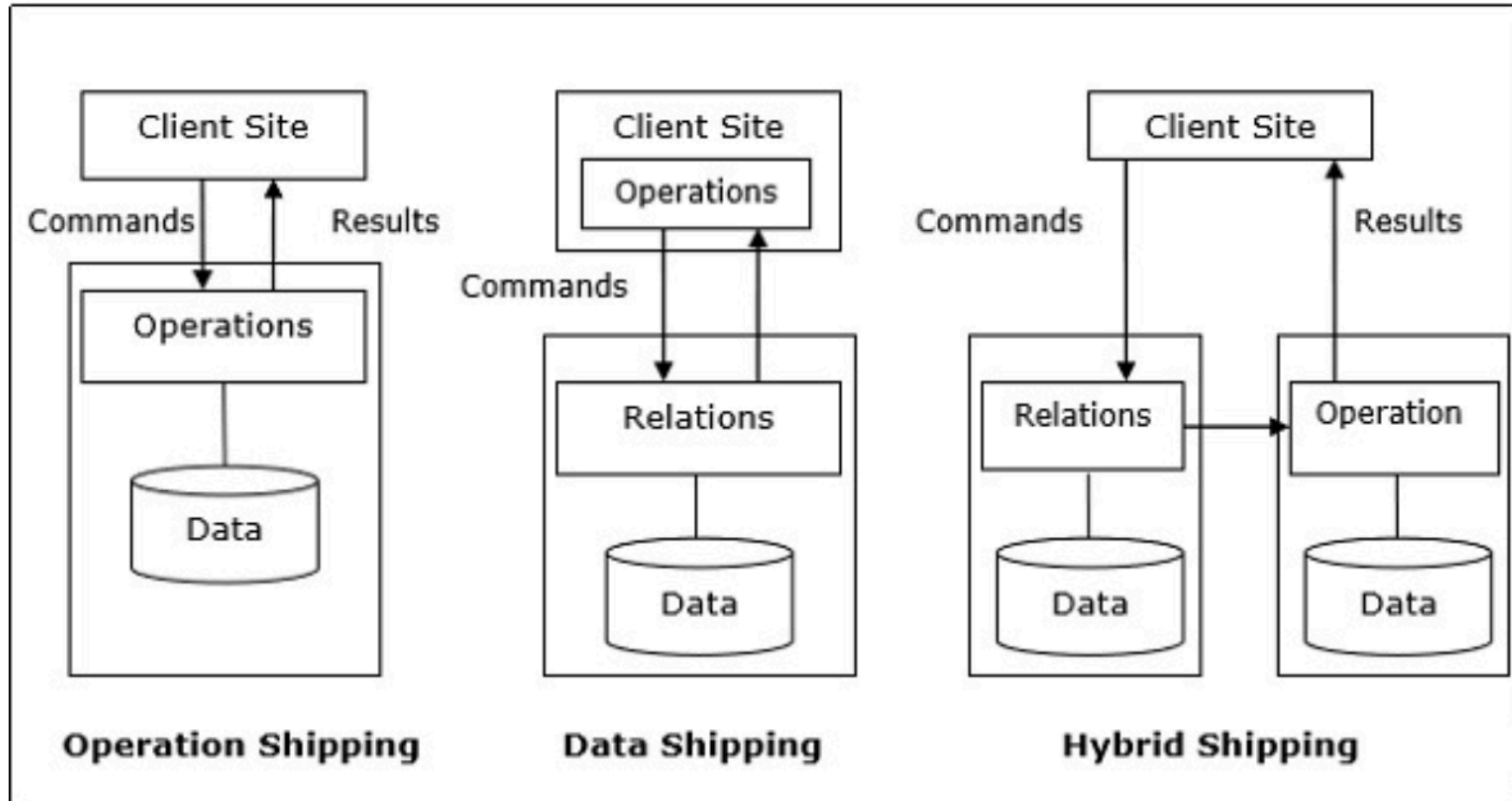
dans une BD distribuée

Utilisation optimale des
ressources dans le SD

- Un système distribué dispose d'un certain nombre de serveurs de base de données dans les divers sites pour effectuer les opérations relatives à une requête
- Les approches d'utilisation optimale de ressources sont:
 - Operation shipping
 - Data shipping
 - Hybrid shipping

Optimisation de requêtes dans une BD distribuée

Utilisation optimale des
ressources dans le SD



Operation shipping

L'opération est exécutée sur le site où les données sont stockées et non sur le site client. Les résultats sont ensuite transférés sur le site client. Ceci est approprié pour les opérations où les opérandes sont disponibles sur le même site

Exemple: Opérations Select et Project

Data shipping

Lors de l'expédition des données, les fragments de données sont transférés vers le serveur de base de données, où les opérations sont exécutées. Ceci est utilisé dans les opérations où les opérandes sont distribués sur différents sites. Ceci est également approprié dans les systèmes où les coûts en communication sont bas et où les processeurs locaux sont beaucoup plus lents que le serveur client

Exemple: Opérations Select et Project

Hybrid shipping

Il s'agit d'une combinaison de data/opération shipping. Ici, les fragments de données sont transférés aux processeurs à grande vitesse, où l'opération est exécutée. Les résultats sont ensuite envoyés au site client

Optimisation de requêtes dans une BD distribuée

Requête trading

Dans l'algorithme de requête trading pour les systèmes de bases de données distribuées, le site de contrôle/client pour une requête distribuée est appelé le **demandeur** et les sites d'exécution des requêtes locales sont appelés **traders**. Le **demandeur** formule un certain nombre d'alternatives pour choisir les **traders** et reconstruire les résultats globaux. L'objectif du **demandeur** est d'atteindre le coût optimal.

Optimisation de requêtes dans une BD distribuée

Requête trading

L'algorithme commence avec le **demandeur** qui attribue des sous-requêtes aux sites du **trader**. Le plan optimal est créé à partir des plans de requête optimisés locaux proposés par les **traders**, combinés au coût de la communication pour reconstruire le résultat final. Une fois que le plan optimal global est formulé, la requête est exécutée.

Optimisation de requêtes dans une BD distribuée

Réduction de l'espace
de solution de la requête

Une solution optimale implique généralement une réduction de l'espace de la solution, de sorte que le coût des requêtes et du transfert de données est réduit. Cela peut être réalisé grâce à un ensemble de règles heuristiques, tout comme les heuristiques dans les systèmes centralisés.

Optimisation de requêtes dans une BD distribuée

Réduction de l'espace
de solution de la requête

- Voici quelques règles
 - Effectuer les opérations de sélection et de projection le plus tôt possible pour réduire le flux de données sur le réseau de communication
 - Simplifier les opérations sur les fragments horizontaux en éliminant les conditions de sélection non pertinentes pour un site particulier
 - Dans le cas d'opérations de jointure et d'union comprenant des fragments situés sur plusieurs sites, transférer les données fragmentées vers le site où la plupart des données sont présentes et effectuer l'opération à cet endroit

Optimisation de requêtes

dans une BD distribuée

Réduction de l'espace
de solution de la requête

- Utilisez l'opération de semi-jointure pour qualifier les n-uplets à joindre. Cela réduit la quantité de transfert de données, ce qui réduit les coûts de communication
- Fusionnez les feuilles et les sous-arbres communs dans une arborescence de requêtes distribuée