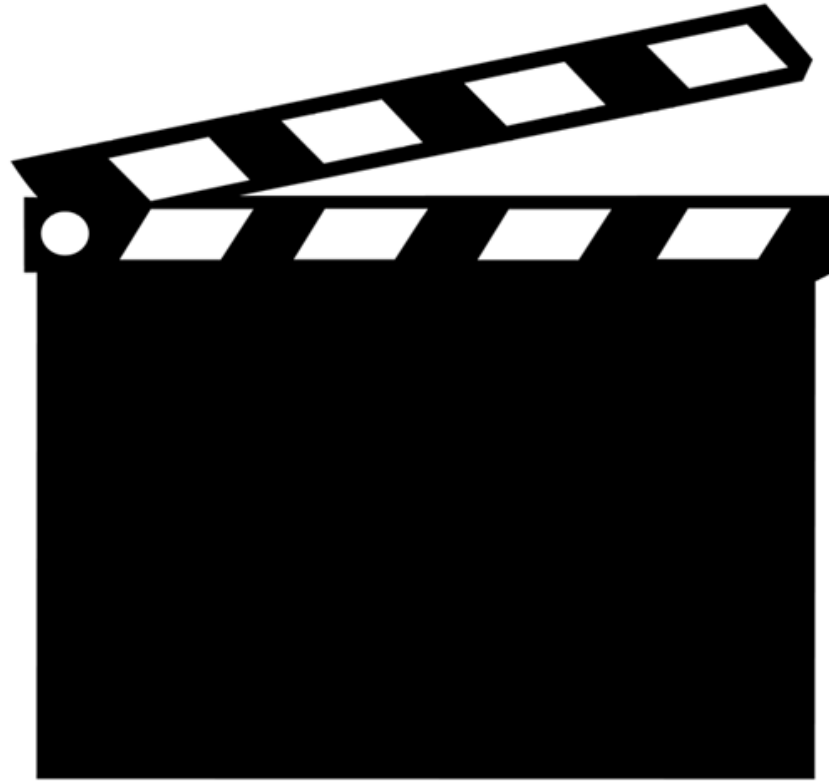


C018SA-W6-S5



SEMAINE 6 : Bases de données distribuées

1. Introduction
2. Différentes architectures
3. Fragmentation
4. Optimisation de requête
- 5. Réplication**
6. Concurrency
7. Conclusion : cinq tendances

Réplication de données : Fiabilité

Faire qu'il y ait plusieurs copies de la même donnée

Si une machine n'est plus accessible, on peut quand même accéder à la donnée

Si une machine perd une donnée, la donnée existe encore

Réplication de données : Performance

Rendre les données plus disponibles

- Réplication d'une relation
- Matérialisation d'une vue

Economiser des communications

Réplication de données : trade-off

Trade-off requêtes vs. mises-à-jour

- Classique en BD

Avantage : Requête

- Interroger une des copies
- Interroger une vue

Désavantage : Mise-à-jour

- Mise-à-jour d'une copie – propager aux autres
- Mise-à-jour d'une vue – traduire en MàJ des relations de bases – ambiguïté

Les copies peuvent ralentir les requêtes (verrous)

Exemple de réplication

Base de données

- Relation Employé E(enum, nom, site, salaire,...)
- Un employé appartient à deux sites en moyennes
- 10 sites d'à peu près la même taille
- 4 à Paris, 4 à Lyon, 2 à Marseille

Charge de travail (*workload*)

- 90% des requêtes à Paris/Lyon/Marseille sont sur des employés locaux

Réplication possible

- 3 SGBD avec les employés locaux

Performance

- Relation Employé E(enum, nom, site, salaire,...)
- Un employé n'appartient qu'à un seul site
- 40 000 employés à Paris et Lyon, 20 000 à Marseille
- 90% des requêtes sur des employés locaux
- Accès disque = 1 unité & Accès réseau = 10 unités
- Sans fragmentation 6.4 unités
- Avec fragmentation 1.9 unités
- Les données répliquées à P/L/M 1 unité

Choisir la réplication

Que doit-on répliquer et où ?

- Problème complexe d'optimisation

Utilise une approche « gloutonne »

Faire jusqu'au point fixe

Pour chaque réplication d'un fragment sur un site

- Quel est le gain
- Quel est le coût

Répliquer le fragment tel que

- $(\text{gain} > \text{coût})$ et $(\text{gain} - \text{coût})$ maximal

Merci

Serge Abiteboul

