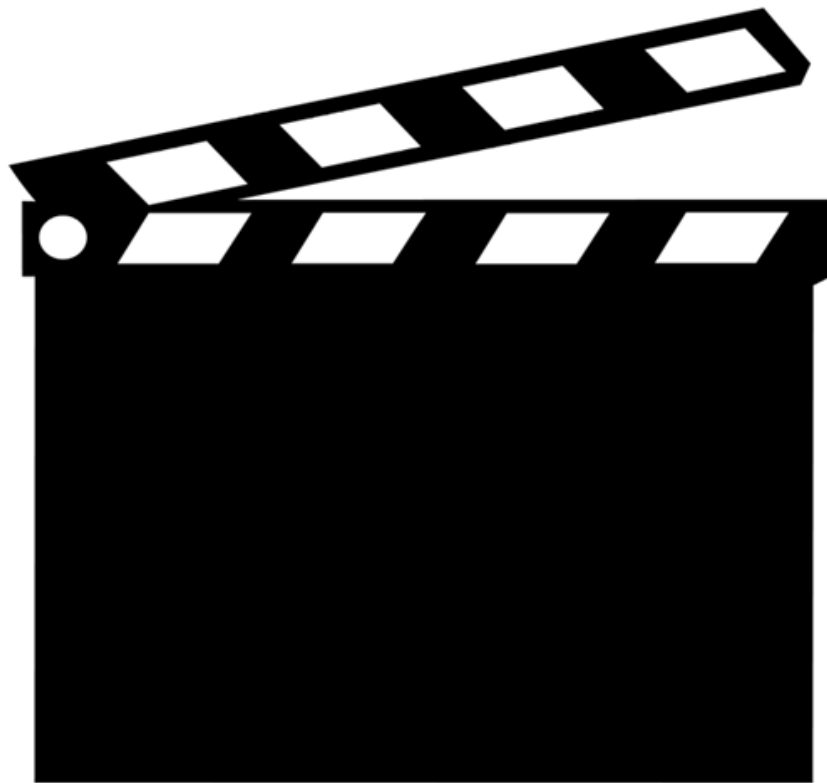


**C018SA-W6-S4**



# SEMAINE 6 : Bases de données distribuées

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

# Que faut-il optimiser ?

## Bases de données classiques

- I/O : Accès disque (millisec)
  - >> Opération du processeur (microsec)

## Bases de données distribuées

- Communication (jusqu'à 1sec sur Internet)
  - >> Accès disque (millisec)

## Attention : basé sur des communications lentes

- par ex, kilo-octets par second
- LAN : bande passante
  - même ordre de grandeur qu'un disque

# Quel est le problème ? Exemple

$E1@site1 = \sigma_{site=Paris}(E)$   $P1@site2 = P \times E1$

$E2@site3 = \sigma_{site \neq Paris}(E)$   $P2@site4 = P \times E2$

- Une requête arrive à site5

**select enom**

**from E, P**

**where E.enum = P.enum**

**and projet = "info"**

$Q = \Pi_{enom} \left( \sigma_{projet="info"}(E \times P) \right)$

$= \Pi_{enom} \left( \left[ \sigma_{projet=info}(E1) \times P1 \right] \cup \left[ \sigma_{projet=info}(E2) \times P2 \right] \right)$

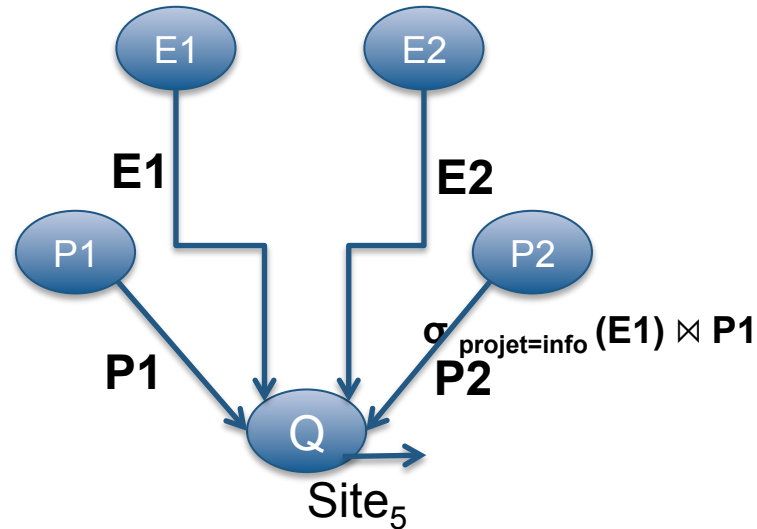
# Plusieurs plans possibles : centralisation brutale

$$E1@site1 = \sigma_{site=Paris}(E)$$

$$E2@site3 = \sigma_{site \neq Paris}(E)$$

$$P1@site2 = P \times E1$$

$$P2@site4 = P \times E2$$



# Plusieurs plans possibles : solution optimisée

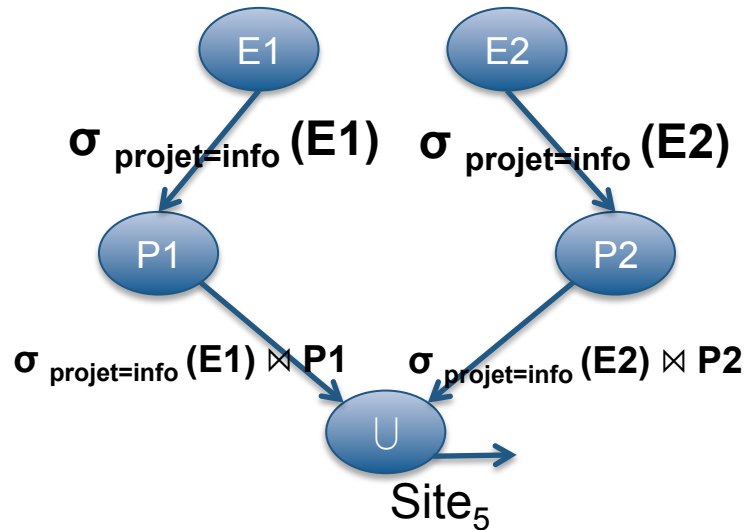
$$E1@site1 = \sigma_{site=Paris}(E)$$

$$E2@site3 = \sigma_{site \neq Paris}(E)$$

$$P1@site2 = P \times E1$$

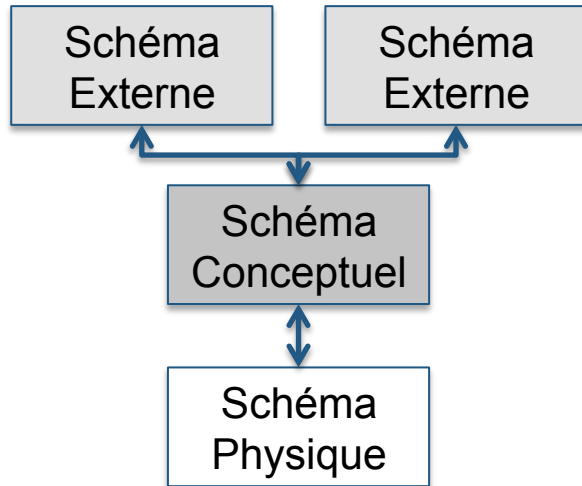
$$P2@site4 = P \times E2$$

Optimisation

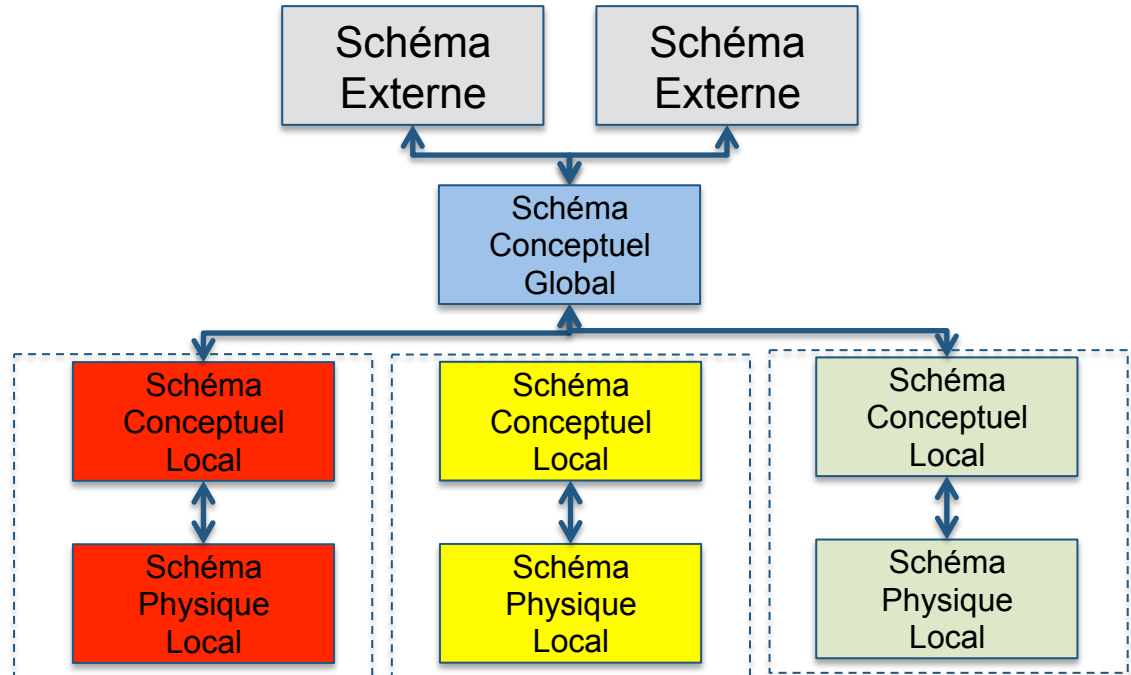


# Architecture logique : intégration de données

- BD centralisé – 3 niveaux



- BD distribuée – 4 niveaux



# Séparer les problèmes pour simplifier

## Analyse syntaxique

- SQL Q on SCG  $\Rightarrow$  requête algébrique H sur SCG

## Localisation

- requête algébrique H sur SCG  $\Rightarrow$  requête algébrique  $G(H_1, H_2)$   
avec  $H_1, H_2$  sur  $SCL_1, SCL_2$

## Optimisation globale

- Minimise globalement les communications
- Obtient  $G(H_1, H_2)$  en  $G'(H_1', H_2')$   
avec  $H_1', H_2'$  sur  $SCL_1, SCL_2$

## Optimisation locale

- Minimise localement les I/O (et aussi les calculs)
- Localement chaque  $H_i'$  est optimisé en  $H_i''$
- On évalue

$$G'(H_1'', H_2'') = G'(H_1', H_2') = H = Q$$



# Merci

Serge Abiteboul

