

## TD n°2

### Requêtes réparties

#### Exercice 1

Soit la BDR composée des relations suivantes :

$\begin{cases} P \text{ (NP, NOMP, MADE\_IN, COULEUR, POIDS)} \\ U \text{ (NU, VILLEU, NOMU)} \\ F \text{ (NF, NOMF, VILLEF, ADRESSE, PAYS, COEF)} \\ PUF \text{ (NP, NU, NF, DATE, QUANTITE)} \end{cases}$

Les relations U et PUF sont sur le site A

Les relations F et P sont sur le site B

Le réseau reliant les deux sites A et B a les caractéristiques techniques suivantes :

- temps d'accès d'un site à un autre : 0,5 seconde
- débit de transmission : 1 000 octets / secondes
- un message est donc transféré en  $(0,5 + \text{Nb-d'octets-du-message} / 1000)$  secondes.

Soit la requête R sur le site B : "Donner les numéros et les noms des fournisseurs qui ont livré un produit italien à une usine située à Lausanne"

$R = \Pi(\text{NOMF, NF}) (\sigma(\text{VILLEU} = \text{"Lausanne"})U * \text{PUF} * \sigma(\text{MADE\_IN} = \text{"Italy"})P * F)$   
où l'opérateur \* est la jointure.

Soient les deux stratégies d'exécution envisagées :

**S1** : Sélectionner les produits italiens sur B. Pour chacun de ces enregistrements, interroger A pour savoir si le produit a été livré dans une usine de Lausanne et par qui. La réponse contient les numéros (NF) des fournisseurs qui livrent ce produit à Lausanne.

**S2** : Sélectionner les numéros (NP) des produits italiens sur B. Transmettre ces numéros au site A en lui demandant qu'il renvoie les numéros des fournisseurs qui ont livré un de ces produits à une usine de Lausanne. Le site A transmet un ensemble de numéros de fournisseurs à B.

Donne le coût en communication pour les deux stratégies précédentes en considérant les données numériques suivantes :

1. F contient 100 fournisseurs.
2. P contient 500 produits italiens, qui concernent 800 livraisons à Lausanne.
3. Les numéros (NP et NF) sont codés sur 2 octets.

#### Exercice 2

Soit **D1**(a, b) sur le site S1, **D2**(a, c) sur S2. La requête  $R1 = D1 \bowtie D2$  est posée sur S3. Le schéma du résultat de la requête est (a, b, c). Les attributs sont des entiers positifs. La distribution des valeurs des attributs est uniforme. La taille, en octets, des attributs *a*, *b*, *c* est respectivement 100, 400 et 3900 octets.

La sélectivité d'une jointure, notée *s*, est telle que

$$\text{card}(D1 \bowtie D2) = s * \text{card}(D1) * \text{card}(D2)$$

On note  $T(D)$  la taille d'une relation D, exprimée en milliards (109) d'octets. On a les valeurs suivantes :

$$T(D1) = 10 \quad T(D2) = 20 \quad s = 0,001\% = 10^{-5}$$

On considère les plans d'exécution suivants :

P1 : transférer D1 et D2 sur S3 puis traiter la jointure sur S3.

P2 : transférer D1 sur S2, puis traiter la jointure sur S2, puis transférer le résultat sur S3.

P3 : transférer D2 sur S1, puis traiter la jointure sur S1, puis transférer le résultat sur S3.

#### Question 1

Le coût d'un plan est la somme des transferts de données mesurés en milliards d'octets.

- a) Quelle est la cardinalité du résultat de la requête R1 ?
- b) Combien vaut  $T(R1)$  ?
- c) Pour chaque plan P1 à P3, donner son coût.

#### Question 2

Pour quelle sélectivité le plan P3 serait meilleur que P1 ?