POLYTECHNIQUE MONTRÉAL

Département de génie informatique et génie logiciel

Cours INF8480: Systèmes répartis et infonuagique (Automne 2021)
3 crédits (3-1.5-4.5)

CORRIGÉ DE L'EXAMEN FINAL

DATE: Mardi le 20 décembre 2022

HEURE: 9h30 à 12h00

DUREE: 2H30

NOTE: Aucune documentation permise sauf un aide-memoire, préparé par l'étudiant, qui consiste en une feuille de format lettre manuscrite recto verso, calculatrice non programmable permise

Ce questionnaire comprend 5 questions pour 20 points

Question 1 (4 points)

a) Le service de noms de domaines (DNS) et le service de répertoire de noms par le protocole LDAP sont deux services de noms fréquemment utilisés. Quelles sont les informations qui sont typiquement placées dans chacun, dans les différentes organisations comme Polytechnique? Dans le protocole LDAP, lors d'une recherche, un temps de traitement et un volume maximal de données sont spécifiés, ce qui n'est pas le cas pour le DNS. Quelle est l'utilité de ces valeurs maximales? Pourquoi ce n'est pas aussi utile pour le DNS? **(2 points)**

Le service de noms de domaines permet essentiellement de convertir les noms de domaines en adresses IP. Il permet aussi de faire la conversion inverse (nom pour une adresse IP). Quelques autres informations peuvent aussi être contenues comme l'adresse du serveur de courriel pour un domaine ou le type d'un ordinateur. LDAP contient la plupart des autres informations requises pour l'accès à un système, comme les noms d'usagers et leur mot de passe (encrypté), les groupes, les privilèges de chaque usager, la localisation du répertoire maison de chaque usager, et potentiellement beaucoup d'autres informations. Etant donné que LDAP permet de mettre toutes sortes d'informations, certaines requêtes pourraient retourner un grand volume d'information ou prendre beaucoup de temps à chercher dans une grande arborescence, même si après filtrage le volume retourné est petit. Dans ce contexte, mettre un quota sur le temps et sur le volume est utile. Avec le DNS, il n'y a pas de recherche dans une arborescence aussi étoffée, ni de filtre, et le résultat retourné est normalement très court, une adresse IP de quelques octets. C'est donc peu utile d'avoir un quota sur le temps ou le volume.

b) Un serveur DNS local s'exécute sur votre poste de travail afin de maintenir un cache local. En effet, de nombreux processus sur votre poste de travail font des requêtes DNS qui peuvent souvent se répéter. Lorsque ce serveur local reçoit une requête, il prend 3ms de coeur de CPU et peut la servir dans 40% des cas. Autrement, il doit en plus faire une demande récursive au serveur DNS de votre fournisseur Internet, ce qui ajoute le temps de cette requête récursive à son temps de réponse. Le serveur DNS de votre fournisseur répond en 10ms dans 65% des cas et en 20ms dans 35% des cas. Quel est le nombre maximal de requêtes par seconde que peut soutenir le serveur local, s'il n'utilise qu'un seul fil d'exécution? S'il utilise de nombreux fils d'exécution et que 2 coeurs de CPU sont dédiés à ce processus serveur local? **(2 points)**

Dans tous les cas, une requête prend 3ms de coeur de CPU. Dans 0.6 x 0.65 des cas, 10ms s'ajoutent, alors que dans 0.6 x 0.35 des cas, l'ajout est de 20ms. Le temps moyen est donc 3ms + 0.6 x 0.65 x 10ms + 0.6 x 0.35 x 20ms = 11.1ms. Avec un seul fil d'exécution, on peut servir 1000ms/s / 11.1ms/r = 90.09r/s. Si plusieurs fils d'exécution et deux coeurs sont disponibles, nous avons besoin de 3ms de coeur de CPU par requête, il est donc possible de servir 1000ms/s / 3ms/r = 333.33r/s pour un coeur, soit le double ou 666.66r/s avec deux coeurs.

Question 2 (5 points)

a) Un ordinateur A envoie un message à B à 11h11m11.150s pour obtenir le temps et reçoit une réponse à 11h11m11.600s, ces deux temps étant mesurés avec l'horloge de A. L'ordinateur B reçoit la requête de A à 11h11m05.200s et retourne sa réponse à A à 11h11m05.500s, ces deux temps étant mesurés avec l'horloge de B. Quel est le décalage à appliquer sur A? Quel est l'intervalle d'incertitude associé? **(2 points)**

a = 11h11m05.200s - 11h11m11.150s = -5.950s
b = 11h11m05.500s - 11h11m11.600s = -6.100s
Ajustement = (a+b)/2 = (-5.950s - 6.100s)/2 = -6.025s
Précision = (a-b)/2 = (-5.950s + 6.100s)/2 = 0.075s

Le décalage à appliquer à A est de -6.025s et l'incertitude est de +/- 0.075s.

b) Un groupe de 25 processus qui communiquent par message sont connectés en anneau. L'algorithme d'élection en anneau est utilisé. Quel est le nombre minimal de messages requis pour compléter une élection? Le nombre maximal? **(2 points)**

Le meilleur cas est lorsque le processus de plus haute priorité amorce l'élection. Après 25 messages, il saura qu'il est élu. Ensuite, après un autre 25 messages, il saura que tous ont appris la nouvelle, pour un total de 50 messages. Le cas le pire est si le processus suivant de celui de plus haute priorité amorce l'élection. Au bout de 24 messages,

celui de haute priorité recevra le message d'élection. Après 25 autres messages il saura qu'il est élu. Finalement, 25 autres messages complèteront la propagation de la nouvelle, pour un total de 74 messages.

c) L'algorithme de l'élection hiérarchique permet de réaliser efficacement une élection. Néanmoins, des algorithmes beaucoup plus complexes comme Paxos et Raft ont été développés pour réaliser des élections. Quels sont les problèmes de l'algorithme de l'élection hiérarchique que l'algorithme Raft permet de pallier? **(1 point)**

Le principal problème de cet algorithme est qu'il n'a pas de notion de quorum ou de majorité absolue. En cas de partition de réseau, il est donc possible d'avoir deux processus qui sont élus, un dans chacune de deux partitions. Pour plusieurs applications comme l'exclusion mutuelle ou les transactions, ceci n'est pas acceptable. Raft ne peut élire un processus que si une majorité absolue accepte sa candidature. Ceci empêche l'élection de plus d'un processus. En plus, Raft est mieux conçu pour des cas de processus qui tombent en panne pendant l'élection, et pour empêcher que tous les processus ne déclenchent une élection en même temps.

Question 3 (4 points)

a) Les transactions T, U, V et W s'exécutent en même temps et leurs opérations de lecture et d'écriture sur des variables (a, b, c, d, e, f, g) sont entrelacées. Les lectures d'une transaction sont effectuées sur les versions courantes des variables, et les écritures d'une transaction sont effectuées sur une version provisoire des variables pour la transaction. Lorsque la transaction se termine et est acceptée, la version provisoire des variables écrites par la transaction devient la version courante. Une validation de la cohérence par contrôle optimiste de la concurrence est effectuée pour accepter ou non chaque transaction. Il faut tenir compte des transactions précédentes qui ont été validées (et ignorer celles qui ne l'ont pas été) pour savoir si chacune des transactions est acceptée ou non. Lesquelles des transactions T, U, V et W pourraient être validées si une validation en avançant était utilisée pour vérifier la cohérence des transactions? Pour chaque transaction non validée, donnez la ou les variables en conflit. **(2 points)**

```
 1 T: Début                    13 T: Compléter
 2 U: Début                    14 W: Read(f)
 3 T: Read(a)                  15 U: Write(c)
 4 T: Read(b)                  16 U: Write(d)
 5 U: Read(c)                  17 U: Compléter
 6 U: Read(d)                  18 V: Read(b)
 7 V: Début                    19 W: Write(a)
 8 V: Read(e)                  20 W: Write(b)
 9 T: Write(b)                 21 W: Compléter
10 T: Write(f)                 22 V: Read(g)
11 W: Début                    23 V: Write(e)
12 W: Read(e)                  24 V: Compléter
```

Au moment de compléter T, il faut vérifier ce que T a écrit (b, f) versus les variables lues par U (c, d), V (e) et W (e). Il n'y a pas d'intersection et T peut compléter. Au moment de compléter U, il faut vérifier ce que U a écrit (c, d) versus les variables lues par V (e) et W (e, f). Il n'y a pas d'intersection et U peut compléter. Au moment de compléter W, il faut vérifier ce que W a écrit (a, b) versus les variables lues par V (b, e). Il y a intersection et W est annulée. La dernière transaction, V, peut nécessairement compléter, puisqu'il n'y a plus d'intersection possible.

b) Une transaction distribuée T, sur 4 serveurs (s1, s2, s3, s4), effectue des opérations sur les variables (a, b, c, d). Chaque variable est répliquée sur deux serveurs, et est dénotée a1 pour la réplique de la variable a qui réside sur le serveur s1. Les opérations effectuées sont: read a1; read b2; read c3; read d4; write a1; write a2; write b2; write b3; write d4; write d1. Ensuite, la transaction répartie est commise en utilisant le protocole de fin de transaction atomique à deux phases. Quelles seront les entrées ajoutées au journal de chacun des serveurs s1, s2, s3 et s4 en lien avec cette transaction T? **(2 points)**

Pour s1: P1: écrire a1; P2 écrire d1; P3 Préparer T(P1, P2) P0; P4 Compléter T, P3;

Pour s2: P1: écrire a2; P2 écrire b2; P3 Préparer T(P1, P2) P0; P4 Compléter T, P3;

Pour s3: P1: écrire b3; P2 Préparer T(P1) P0; P3 Compléter T, P2;

Pour s4: P1: écrire d4; P2 Préparer T(P1) P0; P3 Compléter T, P2;

Question 4 (4 points)

a) Un service de base de données réparti est offert par 3 serveurs redondants. Au moins un serveur doit être disponible pour que le service soit disponible. Chaque serveur est constitué d'un boîtier et son électronique, avec une probabilité de disponibilité de 0.65, ainsi que d'un ensemble de disques en RAID, 5 disques dont au moins 3 doivent être fonctionnels. La probabilité d'être fonctionnel pour un disque est de 0.75. Quelle est la probabilité qu'un ensemble de disques en RAID soit fonctionnel? Un serveur de base de données? Le service de base de données réparti? **(2 points)**

L'ensemble de disques RAID a une probabilité de fonctionner de $0.75^5 = 0.237304688$ *(5 disques fonctionnels) plus* $5!/((5-4)!4!) \times 0.75^4 \times (1-0.75)^{5-4} = 0.395507813$ *(exactement 4 disques fonctionnels) plus* $5!/((5-3)!3!) \times 0.75^3 \times (1-0.75)^{5-3} = 0.263671875$ *(exactement 3 disques fonctionnels), pour un total de* $0.237304688 + 0.395507813 + 0.263671875 = 0.896484376$. *Un serveur est opérationnel si le boîtier est fonctionnel de même que son ensemble de disques RAID, ce qui donne* $0.65 \times 0.896484376 = 0.582714844$. *Le service sera disponible sauf si les 3 serveurs sont en panne, une probabilité de* $1-(1-0.582714844)^3 = 0.927339429$.

b) Considérez la base de données créée avec les commandes suivantes, sur un serveur Postgres configuré de la même manière que lors du travail pratique TP5. Après les commandes pour créer la table ops2, et insérer les valeurs initiales, deux consoles sont utilisées pour entrer deux transactions concurrentes (montrées comme Transaction 1 et Transaction 2). La transaction 2 demande d'imprimer `sum(amount)` à deux reprises. Quelles sont les valeurs imprimées qui ont ici été remplacées par XXXX et YYYY dans la retranscription fournie? **(2 points)**

```
postgres=#CREATE DATABASE bank;
postgres=# \connect bank;
bank=# CREATE TABLE ops2 (id int, amount float, PRIMARY KEY (id));
bank=# INSERT INTO ops2 VALUES (1,-100);
bank=# INSERT INTO ops2 VALUES (2,+150);
bank=# INSERT INTO ops2 VALUES (3,-22.2);

Temps Transaction 1    Transaction 2

1                      bank=# BEGIN TRANSACTION ISOLATION LEVEL REPEATABLE READ;
2                      bank=#  SELECT sum(amount) FROM ops2;
                       sum
                       ------
                       XXXX

3    bank=# BEGIN;
4    bank=# INSERT into ops2 VALUES (4,150);
5    bank=# COMMIT;

6                      bank=#  SELECT sum(amount) FROM ops2;
                       sum
                       ------
                       YYYY
```

Dans le mode d'isolation `REPEATABLE READ` *utilisé par la transaction 2, lorsque la même valeur est lue plus d'une fois dans la même transaction, la même valeur est obtenue à chaque fois. La mise à jour, effectuée aux temps 3 à 5 par l'autre transaction, ne peut donc pas être prise en compte dans ce mode. Le résultat pour XXXX est donc -100 + 150 - 22.2 = 27.8, et il est le même pour YYYY.*

Question 5 (3 points)

Vous devez planifier un nouveau centre de données et comparer différents scénarios. Vous avez déjà prévu examiner la performance des systèmes (et les revenus qu'on peut en tirer), le coût des ordinateurs, le coût du bâtiment ainsi que les coûts d'opération et de renouvellement des équipements, afin de déterminer le projet le plus rentable sur la durée de vie anticipée. Devez-vous comme ingénieur aussi vérifier les aspects du développement durable de ce projet? i) Quelles sont les lois applicables? ii) Quelles sont les différentes phases du cycle de vie? iii) Quels sont les quatre différents types d'impact sur l'environnement et ceux les plus importants typiquement pour un centre de données dans chaque phase? **(3 points)**

L'ingénieur doit tenir compte des conséquences de l'exécution de ses travaux sur l'environnement dans une perspective de développement durable. i) Ceci est prescrit par la loi canadienne de 2008 sur le développement durable, la loi québécoise

de 2006 sur le développement durable, et le code de déontologie de l'Ordre des Ingénieurs du Québec. La procédure d'évaluation environnementale du BAPE ne s'applique normalement pas à un centre de données. ii) Les phases du cycle de vie sont la fabrication (construction du site et fabrication des équipements), le transport (des matériaux pour la construction, et des équipements), l'opération du site (entretien, alimentation électrique...) et finalement la fin de vie (la démolition ou conversion du bâtiment, la restauration du site, et le recyclage ou la mise aux rebuts des matériaux et équipements). iii) Pour toutes les activités qui se retrouvent dans ces différentes phases, il faut examiner l'impact sur la santé humaine, sur l'écologie, sur les changements climatiques et sur l'appauvrissement des ressources. Pour un centre de données typique, une grande partie de l'impact est reliée à la consommation d'énergie pendant l'opération de l'équipement informatique et de la climatisation (réchauffement climatique et appauvrissement des ressources, santé humaine et écologie aussi si charbon). Un autre impact non négligeable est la fabrication des équipement informatiques et électriques, en particulier le raffinage de l'or et du cuivre requis pour ces équipements (santé humaine et écologie). Etant donné l'importance de la consommation électrique, une source d'énergie propre, et un climat froid qui ne requiert pas de climatisation, peuvent diminuer considérablement l'impact négatif d'un centre de données.

Le professeur: Michel Dagenais