

TD2-M2: Workloads

S. Zertal

Caractérisation d'une trace

On considère une trace d'accès à un système de stockage (voir trace.log en fichier joint). Chaque ligne correspond à une requête de lecture ou d'écriture et respecte le format suivant :

<estampille> <type> <taille> <date> <adr>

estampille	numéro de la requête [0..max-1]
type	0:lecture , 1:écriture
adr	adresse de la requête dans l'espace global (num bloc)
taille	taille en nombre de blocs de 4KB chacun
date	date d'arrivée (ms)

A partir de ce fichier de trace, déterminez le profil des accès à ce système en répondant aux questions suivantes :

1. Les accès sont-ils orientés lecture ou écriture ? Justifiez votre réponse.
2. Les accès sont-ils courts ou longs ? Justifiez votre réponse en vous basant sur la taille moyenne, détaillez en analysant sa variabilité. Y a-t-il un lien avec le type des accès ?
3. Tracez l'histogramme montrant la répartition des tailles de requêtes de cette trace. Qu'en déduisez vous ?
4. Calculez les quartiles pour la taille des requêtes tout types confondus puis en distinguant les types. Calculez l'écart entre le 1er et le 3ème quartile. Que déduisez vous de cet écart ? Correspond il aux résultats précédents ?
5. Quelle est la médiane de la taille des requêtes ? comparez la à la taille moyenne, qu'en déduisez vous ?
6. Analysez les dates d'arrivées des requêtes tout types confondus et dites s'il y a un effet rafale. A refaire en distinguant les lectures des écritures.
7. Analysez les adresses des requêtes, les accès sont-ils uniformément distribués sur l'espace ? ou il y a t-il des hot spots ?

Génération d'une trace

On aimerait analyser l'effet d'une répartition spatiale non uniforme des accès au système de stockage de l'exercice précédent, sur ses performances en terme de temps de réponse. Pour cela, on doit développer un générateur de requêtes intégrant l'option hot spot avec comme paramètre :

- S_H : ($Size_{Hot-Spot}$) la taille de la zone active ;le pourcentage qu'elle représente au sein de l'espace global)
 - I_H : ($Intensity_{Hot-Spot}$) le facteur de multiplication d'accès à cette zone ; le facteur pour la zone restante étant 1)
1. Implémentez en C (ou C++ ou python), la programme qui réalise cette génération.
 2. Faites un test pour 10K requêtes de lecture/écriture sur un espace de 144GB dont 10% est concerné par 90% des accès.
 3. Tracez le graphe représentant le nombre d'accès en fonction de l'emplacement pour illustrer la distribution non uniforme des accès (hot spots).

La sortie sera un fichier "traceHspot" avec le format : estampille adresse **A**
rendre :

1. Le code source
2. Un readme qui détaillera le mode d'exécution de votre programme
3. La trace générée compressée
4. Le graphe correspondant en format pdf