

Concaténation de fichiers SAS

Exposé du problème

Une procédure stockée produit 20 fichiers intermédiaires qui doivent être regroupés en un seul fichier de sortie. Pour ce faire, l'instruction SET est utilisée. Présentement, cette étape prend 1 heure sur les 3 que prend l'exécution de la totalité de la procédure stockée. Il faudrait optimiser cette étape afin de réduire le temps d'exécution du traitement.

Caractéristiques des données

Chacun des fichiers contient environ 19 millions de lignes et 83 colonnes.

Chacun des fichiers est compressé en binaire.

Le fichier de sortie contient 380 millions de lignes et 83 colonnes et est compressé en binaire. Il occupe 55 G.

L'exécution se fait sur SAS 9.4 avec SAS EG 6.1.

Utilisation du disque physique (local) sur le serveur SAS au lieu du WORK.

Pistes explorées

Mise en mémoire des 20 tables produites par la procédure stockée.

Façon de faire : SASFILE << nom de la table >> open;

Problème : atteinte de la limite de la mémoire du serveur.

Proc Append : qui consiste à créer un data avec la première partition et ensuite ajouter les autres partitions une par une. Plus long que l'instruction SET.

Avec le « multithreading », l'instruction SET faite sur deux processus SAS de 10 partitions chacune, le temps sauvé est infime.

Tests effectués :

```
/* Affecter le disque local du serveur SAS */  
libname W "\\Sdiaatdeg20\AppStock\Point moyen";
```

1. SET de 20 partitions. Le temps obtenu a été de 1 heure.

```
data w.resultat;  
  set w.partition_1  
      w.partition_2  
      w.partition_3  
      w.partition_4  
      w.partition_5
```

```

w.partition_6
w.partition_7
w.partition_8
w.partition_9
w.partition_10
w.partition_11
w.partition_12
w.partition_13
w.partition_14
w.partition_15
w.partition_16
w.partition_17
w.partition_18
w.partition_19
w.partition_20;
run;

```

2. SET de la première partition et PROC APPEND pour les suivantes. Le temps obtenu est d'environ 11 minutes juste pour les 3 premières partitions (trop long)

```

data w.resultat;
    set w.partition_1;
run;

%macro ajouter;

    %do i=2 %to 20;

        proc append base=W.resultat
data=w.partition_&i.;
            run;

    %end;
%mend;
%ajouter;

```

3. Démarrer 2 processus qui concatènent chacun 10 partitions et enfin concaténer le résultat des deux processus. Le temps total obtenu est d'environ 43 min.

```

options sascmd="sas";

signon tachel;
rsubmit process=tachel sysrputsync=yes wait=no;
libname W "\\Sdiaatdeq20\AppStock\Point moyen";
compress=binary;

```

```

data w.resultat_1;
  set w.partition_1
      w.partition_2
      w.partition_3
      w.partition_4
      w.partition_5
      w.partition_6
      w.partition_7
      w.partition_8
      w.partition_9
      w.partition_10;

run;
endrsubmit;

signon tache2;
rsubmit process=tache2 sysrputsync=yes wait=no;
libname W "\\Sdiaatdeq20\AppStock\Point moyen"
compress=binary;
data w.resultat_2;
  set w.partition_11
      w.partition_12
      w.partition_13
      w.partition_14
      w.partition_15
      w.partition_16
      w.partition_17
      w.partition_18
      w.partition_19
      w.partition_20;

run;
endrsubmit;

waitfor ALL tache1 tache2;
signoff tache1;
signoff tache2;

data w.resultat;
  set w.resultat_1 w.resultat_2;
run;

```

Y a-t-il une méthode pour concaténer rapidement des tables SAS ?