



Master d'Informatique

UE INF2245

Calcul haute performance pour le «Big Data»

Sujet d'examen terminal (seconde session)

Durée : 2 heures
Aucun document autorisé

F. Raimbault, N. Courty

juin 2021

Avertissement : Il sera principalement tenu compte dans la notation de la qualité de rédaction, de votre capacité à abstraire une solution et de la rigueur avec laquelle vous la décrivez.

Exercice 1 : CUDA (7 points)

1. Rappelez en quelques lignes les principes de la programmation en cuda : quels avantages ? Dans quels cas de figure peut on vouloir utiliser ce modèle de programmation ? Quels sont les pré-requis ?
2. Soit le programme suivant :

```
import pycuda.autoinit
import pycuda.driver as drv
import numpy

from pycuda.compiler import SourceModule
mod = SourceModule("""
__global__ void my_function(float *dest, float *a, float *b, int size)
{
    const int i = threadIdx.x;
    if (i < size){
        dest[size - i - 1] = 2*a[i] - b[i];
    }
}
""")

my_function = mod.get_function("my_function")

a = numpy.array([0,1,2,3,4,5,6,7,8,9])
b = numpy.array([9,8,7,6,5,4,3,2,1,0])

dest = numpy.zeros_like(a)
my_function(
    drv.Out(dest), drv.In(a), drv.In(b), 10)
    block=(10,1,1), grid=(1,1))

print(dest)
```

Quelle est la sortie de l'exécution de ce programme ? Cette sortie est elle déterministe (toujours le même résultat) ? Justifiez votre réponse.

3. Même question si on change l'appel du noyau par :

```
my_function(
    drv.Out(dest), drv.In(a), drv.In(b), 10)
    block=(5,1,1), grid=(2,1))

print(dest)
```

4. Modifiez le noyau my_function de manière à ce que le programme s'exécute de la même manière pour l'appel précédent que dans la question 2.

Exercice 2 : Frameworks de programmation distribuée pour le Big Data (7 points)

Répondez précisément aux questions suivantes sans mélanger les réponses aux différentes questions.

1. Décrivez le modèle de programmation MapReduce.
2. Décrivez le modèle de programmation Spark.
3. En quoi sont ils comparables ?
4. En quoi ces deux modèles diffèrent ?

Exercice 3 : Modélisation d'un graphe (6 points)

Vous avez à traiter des algorithmes sur des graphes de taille très importante (ex : liens entre tous les membres du réseau social Facebook).

1. En quoi le cadre de conception MapReduce est adapté pour traiter ce genre d'application ?
2. Comment modéliser un graphe afin de pouvoir écrire des algorithmes sur ces données avec MapReduce ?
3. En quoi le cadre de conception Spark est adapté pour traiter ce genre d'application ?
4. Comment modéliser un graphe afin de pouvoir écrire des algorithmes sur ces données avec Spark ?