

Contrôle Continu : ITI534 Big Data

1 Octobre 2019

Durée 1h, tous documents non électroniques autorisés.

Le barème est donné à titre indicatif.

Dans tous les exercices sur la complexité, nous supposons, sauf indication contraire, que les opérations suivantes prennent un temps constant $O(1)$: addition de deux nombres, multiplication de deux nombres, division de deux nombres, soustraction de deux nombres et la racine carrée.

- (3 points) Quelles sont les caractéristiques principales du phénomène “Big data” ? (Ces caractéristiques sont souvent abrégée en VVV.)
- (2 points) Donnez un exemple du problème type “Big Data” ayant une place avant le vingtième siècle. Comment les gens ont-ils géré le traitement de ces données ?
- (3 points) Donnez au moins trois exemples de langages de programmation utilisés pour le traitement de grandes quantités de données.
- (2 points) Pourquoi Spark est plus rapide que Hadoop sur les problèmes de tri ?
- (3 points) Supposons que nous ayons déjà calculé la moyenne

$$M_n = \frac{\sum_{i=1}^n x_i}{n}$$

de n éléments. Expliquer le fonctionnement de l’algorithme rapide de mise à jour de valeur moyenne lorsque nous ajoutons un nouvel élément x_{n+1} à notre échantillon statistique. Quelle est la complexité de l’algorithme ?

- (5 points) Pour estimer l’écart type d’un ensemble $X = \{x_1, x_2, \dots, x_n\}$ ayant n éléments on utilise la formule suivante

$$\sigma(X) = \sqrt{\frac{\sum_{i=1}^n (x_i)^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}$$

Décrivez une méthode rapide de mise à jour d’écart type lorsqu’un nouvel élément s’ajoute dans l’ensemble X . Quelle est la complexité de votre méthode ?

- (2 points) Quel algorithme préféreriez-vous pour faire un traitement de grandes quantités de données ? Pourquoi ?
 - un algorithme avec la complexité $O(n^2)$
 - un algorithme avec la complexité $O(\sqrt{n})$
 - un algorithme avec la complexité $O\left(\frac{n}{1000000}\right)$
 - un algorithme avec la complexité $O\left((\ln(n))^2\right)$
 - un algorithme avec la complexité $O(2^n)$