

Contrôle : Big Data 2025

Durée 1h, tous documents non électroniques autorisés.

Le barème est donné à titre indicatif.

1. (4 points) Pourquoi les questions éthiques sont-elles souvent soulevées lors de la collecte et de l'analyse de données ?
2. (4 points) Y a-t-il une différence entre $O(n^3)$ et $O(10n^2 + n^3)$? Expliquer pourquoi.
3. (4 points) Un algorithme effectue $10 \cdot \sqrt{n} \cdot n^3$ multiplications, où n est la taille de données en entrée. De combien de fois augmente le nombre de multiplications si nous avons quatre fois plus de données en entrée ?
4. (4 points) Qu'est-ce qu'un facteur d'approximation ?
5. (4 points) La *moyenne glissante sur k dernières valeurs* d'une série $x_1, x_2, \dots, x_n$ de n valeurs est défini comme

$$\frac{x_{n-k+1} + x_{n-1} + x_n}{k}.$$

Décrivez la méthode rapide de mise à jour de la moyenne glissante sur k dernières valeurs lorsque un nouvel élément x_{n+1} s'ajoute. Essayez de proposer un algorithme rapide et peu consommateur en mémoire. Quelle est la complexité de votre méthode en espace mémoire et en temps de calcul ?