



Computer Science Department

Examen INF5099 Apprentissage Distribué, 14 Avril 2021

Docteur Paulin MELATAGIA, Docteur Thomas MESSI

**Exercice 1 : Map Reduce (6.5 points)**

1. Rappeler le principe de Map Reduce tel que vu en cours **(0.5 point)**
2. Dites dans quel cas Map Reduce peut-être vu comme un model et dans quel cas il peut-être vu comme une technologie. **(0.25 \* 2 point)**
3. Choisir un algorithme parmi ceux présentés en classe lors des exposés (SVM, Ensemble Learning, K-means, Neural Network, Logistic Regression, PCA) et répondez aux questions suivantes : **(5 points)**
  - (a) Rappeler la version sequentielle de l'algorithme. **(1 point)**
  - (b) Donner une stratégie de parallélisation et justifier en quoi elle permettra de réduire le temps d'exécution. **(1 point)**
  - (c) Réaliser une implémentation de l'algorithme en utilisant Map Reduce et posix thread. (On donnera juste l'essentiel : code des mappers, code des reducers). **(2.5 points)**
  - (d) Rappeler et expliquer la formule du calcul du temps d'exécution d'un programme parallèle d'après la loi d'Amdahl. Qu'en est-il de votre implémentation ? **(1 point)**

**Exercice 2 : DryadLink. (3.5 points)**

1. DryadLink est un modèle ou une technologie ? Justifier votre réponse. **(0.5 point)**
2. Prenez un algorithme d'apprentissage de votre choix et dites comment réaliser sa version parallèle avec DryadLink (sous forme de grandes lignes). **(1.5 points)**
3. Le problème principal de DryadLink est son lien fort avec les technologies de Microsoft. On voudrait résoudre ce problème en réalisant une implémentation plus libre. Après avoir rappeler le fonctionnement de DryadLink, dites-nous comment réaliser un outil équivalent en se servant de Map Reduce. **(1.5 points)**

..... Bon Courage ! .....