CSE 234 Fall 2021 Surprise Quiz 1

INSTRUCTIONS:

1. This quiz is for **10min** and **10pts**. It is open book/notes/electronics.

2. Mark/write your answers using a dark pen or pencil. Ensure your writing is clear and legible.

**Q1. [2pts]** For which of these ML methods will model averaging-based SGD not converge well for distributed execution?

(A) Linear regression (B) Logistic regression (C) Deep neural network (D) Linear SVM

**Q2. [2pts]** Which function in the UDA API roughly corresponds to the role of Map in MapReduce?

(A) Initialize (B) Transition (C) Merge (D) Finalize

**Q3. [2pts]** What is the main type of data partitioning scheme used by XGBoost for distributed execution?

(A) Row-oriented (B) Column-oriented (C) Tile-oriented (D) None of these

**Q4. [4pts]** Briefly and clearly explain 1 advantage and 1 disadvantage of Parameter Server over in-RDBMS ML.
CSE 234 Fall 2021 Surprise Quiz 1

Full Name: ___________________________ PID: ___________________________

INSTRUCTIONS:

1. This quiz is for **10 min** and **10 pts**. It is _open_ book/notes/electronics.

2. Mark/write your answers using a dark pen or pencil. Ensure your writing is clear and legible.

**Q1. [2pts]** What is the main type of data partitioning scheme used by XGBoost for distributed execution?

   (A) Row-oriented    (B) Column-oriented    (C) Tile-oriented    (D) None of these

**Q2. [2pts]** For which of these ML methods will model averaging-based SGD not converge well for distributed execution?

   (A) Linear regression (B) Logistic regression (C) Deep neural network (D) Linear SVM

**Q3. [2pts]** Which function in the UDA API roughly corresponds to the role of Map in MapReduce?

   (A) Initialize    (B) Transition    (C) Merge    (D) Finalize

**Q4. [4pts]** Briefly and clearly explain 1 advantage and 1 disadvantage of ML-on-dataflow systems over Parameter Server.
INSTRUCTIONS:

1. This quiz is for **10min** and **10pts**. It is *open* book/notes/electronics.

2. Mark/write your answers using a dark pen or pencil. Ensure your writing is clear and legible.

**Q1. [2pts]** Which function in the UDA API roughly corresponds to the role of Map in MapReduce?

(A) Initialize (B) Transition (C) Merge (D) Finalize

**Q2. [2pts]** What is the main type of data partitioning scheme used by XGBoost for distributed execution?

(A) Row-oriented (B) Column-oriented (C) Tile-oriented (D) None of these

**Q3. [2pts]** For which of these ML methods will model averaging-based SGD not converge well for distributed execution?

(A) Linear regression (B) Logistic regression (C) Deep neural network (D) Linear SVM

**Q4. [4pts]** Briefly and clearly explain 1 advantage and 1 disadvantage of ML-on-dataflow systems over in-RDBMS ML.