Electrical and Computer Engineering / Software Engineering

CPRE/SE 419 Software Tools for Large Scale Data Analytics

Spring 2023

Homework 2

Due: Tuesday, April 18, 11:59PM

Preamble:
The purpose of this homework is for you to practice a bit of a "lingo" for the paradigms covered in class in the past few weeks. You have 7 problems and you should try to provide concise (and, of course, correct) answers.

Problem Set

1. (12 pts.) Describe briefly the concept of *false cycle* in distributed transaction management

2. (12 pts.) Describe briefly the (difference between) *transformations* and *actions* in Spark.

3. (15 pts.) What are the benefits and drawbacks of the *Majority Protocol* for distributed lock management?

4. (15 pts.) Describe briefly the *CAP theorem.* Provide an example of a NoSQL database from the "AP" spectrum.

5. (15 pts.) Describe briefly the concept of a *polystore*.

6. (12 pts.) Describe briefly the difference between *streaming* algorithms and *online* algorithms.

7. (19 pts.) Consider the following schedule of 3 transactions accessing 3 different data-items:

TR1	TR2	TR3
read(B)		
read(A)		
write(B)		
	write(A)	
	read(C)	
		write(C)
		read(B)
	write(C)	
	write(B)	
write (B)		

- Would you say that this schedule is conflict-serializable (justify your answer)?

What to turnin: Typed solutions are strongly preferred. If, for whatever reason, you are prevented from using any editor, then we may accept hand-written solution – provided that they are legible.

This is assignment can be done in teams of two students (of course, we will honor individual submissions. While one submission per team is enough – please make sure to indicate the names of both team members (or, to explicitly state that this was an individual assignment) in the preamble of the document with your solutions.