

Homework 8: Spark Framework

Q1: What are the features of Spark RDD?

- A. In-memory computation
- B. Lazy evaluation
- C. Fault tolerance
- D. All of the above

Q2: How many Spark Context can be active per JVM?

- A. Two
- B. One
- C. Unlimited

Q3: Which of the following is not a transformation?

- A. `flatMap(func)`
- B. `map(func)`
- C. `reduce(func)`
- D. `filter(func)`

Q4: Which one of the following operations does NOT trigger an eager evaluation?

- A. `take(n)`
- B. `collect()`
- C. `count()`
- D. `join(otherDataset, [numPartitions])`

Q5: Which of the following commands will NOT generate a shuffle of data from each executor across the cluster?

- A. `collect()`
- B. `map(func)`
- C. `repartition(numPartitions)`
- D. `distinct([numPartitions])`