CSP 554 – Big Data Technologies

Fall 2022 – All Sections

Midterm Exam - Sample Questions

Part I – True/False (Circle either T or F) – 1 points each, 15 points total

1. T or F – A horizontally-scaled distributed system has easily managed shared state and calculations versus a vertically-scaled centralized system.

Part II – Multiple Choice (Circle Either a, b, c, or d) – 2 points each, 20 points total

1. Which of the following components of the Hadoop ecosystem performs the function of scheduling tasks and monitoring resource utilization across a cluster:

a. HDFS

c. Hive

b. YARN

d. None of the Above

Part III – Fill in the Blank (Write answer in provided space) – 3 points each, 15 points total

1. A custom Combiner function that is defined for a Map-Reduce job is guaranteed to execute _____ times.

Part IV – Short Answer (Show Points/Results) – 4 points each, 20 points total

1. Given the following steps of a Map-Reduce job, order them in correct sequence that they are executed in: Partition, Combine, Map, Shuffle, Reduce. Which steps are guaranteed to execute? Which are not?

Part V – Long Answer (Show Reasoning/Calculations) – 5 points each, 20 points total

1. Given the following key-value pair results from a Map step of a Map-Reduce job, what would be the resulting key-value pairs that would be the input to a Reduce job after the Shuffle/Sort step: <hello, 1>, <there, 1>, <hello, 1>, <again, 1> ? Would this result be different if a Combiner step was run? If so, provide the resulting key-value pairs in that case.

Part VI – Essay Question (Show Argument/Proof) – 10 points each, 10 points total

1. Outline how an internal Hive table that is partitioned is stored on HDFS, and how a HQL query can take advantage of this to minimize I/O. For an example table consisting of the following fields, which would be the most likely candidates for partitioning? Explain your reasoning.

CustomerID

CustomerFirstName

CustomerLastName

CustomerStreetAddress

CustomerCity

CustomerState

CustomerZipCode

CustomerPhoneNumber