

CSP554—Big Data Technologies

Assignment #8

Worth: 6 points

Due by the start of the next class period

Assignments can be uploaded via the Blackboard portal

Read from (TW)

- Chapter 19 (review one more time)
- Review other Spark materials on the Blackboard

Exercise 1: Read the article “The Lambda and the Kappa” found on our blackboard site in the “Articles” section and answer the following questions using between 1-3 sentences each. Note this, article provides a real-world and critical view of the lambda pattern and some related big data processing patterns:

1. (1 point) Extract-transform-load (ETL) is the process of taking transactional business data (think of data collected about the purchases you make at a grocery store) and converting that data into a format more appropriate for reporting or analytic exploration. What problems was encountering with the ETL process at Twitter (and more generally) that impacted data analytics?
2. (1 point) What example is mentioned about Twitter of a case where the lambda architecture would be appropriate?
3. (2 points) What did Twitter find were the two of the limitations of using the lambda architecture?
4. (1 point) What is the Kappa architecture?
5. (1 point) Apache Beam is one framework that implements a kappa architecture. What is one of the distinguishing features of Apache Beam?