

## CSP554—Big Data Technologies

### Assignment #9

#### Readings

Read Chapters 1, 2 and 3 from our next book: Pramod J. Sadalage and Martin Fowler. 2012. *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Addison-Wesley.(PS)

#### Worth: 5 points + 5 points extra credit

#### **Due by the start of the next class period**

Assignments should be uploaded via the Blackboard portal

Exercise 1) 5 points

Read the article “Real-time stream processing for Big Data” available on the blackboard in the ‘Articles’ section and then answer the following questions:

- a) (1.25 points) What is the Kappa architecture and how does it differ from the lambda architecture?
- b) (1.25 points) What are the advantages and drawbacks of pure streaming versus micro-batch real-time processing systems?
- c) (1.25 points) In few sentences describe the data processing pipeline in Storm.
- d) (1.25 points) How does Spark streaming shift the Spark batch processing approach to work on real-time data streams?

Exercise 2) 5 points extra credit

Follow the document “Instructions for setting up a VM with Kafka” included with this assignment and execute the demo code. Provide enough screen shots to indicate you have completed the document through section 4. Then remember to terminate your VM.