# CS595—Big Data Technologies

## Assignment #1 (Modules 01a & 01b, 10 points + 2 points extra credit)

## Due by the start of the next class period

Assignments are to be uploaded via the Blackboard portal

Note: There may be short quiz questions about the readings or articles (except extra credit) in the class period when they are due.

1. Obtain our texts

* Tom White. 2015. Hadoop: The Definitive Guide (4th ed.). O'Reilly Media, Inc (TW)
* Pramod J. Sadalage and Martin Fowler. 2012. *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Addison-Wesley.(PS)

1. Read from (TW)

* Chapter 1 (note this chapter is also on Blackboard “Free Books and Chapters” so you don’t need to wait for the book to arrive)
* Chapter 3

1. (5 points) Submit very brief answers (or bullet points) to the following questions:

* Describe any prior experience you might with, data mining, machine learning, statistics, data science and big data
* Share any big data interests and personal learning goals for the course
* Indicate if there are additional topics in the scope of the course of special interest to you
* Indicate if you have access to big data technology and data sets, of what nature, and in what industry.
* Do you have any anticipated personal issues such expected absences or other necessary accommodations with course impact? (Of course, these will be held in strictest confidence.)

1. Read article on “Blackboard” in Articles section

* The Parable of Google Flu (just 3 pages!)

1. (5 points) Summarize the main points of the above article and your thoughts (questions you might want to ask the authors, areas where you disagree, other comments)

* No more than about ½ page single spaced
* Submit via blackboard

Extra Credit:

1. Read article on “Blackboard” in Articles section

* Byzantine Fault Tolerant MapReduce

1. (2 points) Summarize the main points of the above article and your thoughts (questions you might want to ask the authors, areas where you disagree, other comments)

* No more than about ½ page single spaced
* Submit via blackboard