**INFS 774 Big Data Analytics**

**Guidelines for Week 2**

Relevant Learning Objectives:

1. Gain an overview of Apache Hadoop and its advantages over traditional large-scale computing systems.
2. Gain an overview of various key technologies in the Hadoop ecosystem.
3. Gain an overview of HDFS and develop an understanding about how data are stored in HDFS

Required activities:

1. Watch Professor Liu’s week 2 lecture videos posted on the course video site. The accompanying slides have been posted under the “week 2” folder.
2. Read chapter 4 “All about Hadoop -the Big Data Lingo” of the book “Understanding Big Data”
3. Read chapter 2 “HDFS” of the book “Hadoop Operation”

Optional activities:

1. Frequently used linux command <http://faculty.tru.ca/nmora/Frequently%20used%20UNIX%20commands.pdf>
2. Frequently used Hadoop hdfs command <https://knpcode.com/hadoop/hdfs/frequently-used-hdfs-commands/>
3. Hadoop use cases

<https://hadoopilluminated.com/hadoop_illuminated/Hadoop_Use_Cases.html>

<https://www.computerworlduk.com/galleries/data/hadoop-distributions-in-enterprise-3641429/>

<https://channels.theinnovationenterprise.com/presentations/hadoop-use-cases-and-how-ebay-operates-its-multi-petabyte-clusters-efficiently>