

# TDT4305 2021 - Assignment 1

## Intro

- 1) For each of the three Vs, answer the following:
  - a) Why is this property showing up in the Big Data era and not previously?
  - b) What challenges does this property give for traditional RDBMSs?
- 2) Given Big Data from a field of your choice, how could you use it to create value for yourself or others? E.g. Big Data in medicine can be used for diagnosing diseases based on patients' health history.
- 3) What are the challenges in ensuring the trustworthiness of Big Data?

## MapReduce and HDFS

- 1) Solve the following exercises from the main course book MMDS:
  1. 2.2.1 (a) and (b)
  2. 2.3.1 (a), (b), and (c)
  3. 2.3.2
- 2) What is the role of the DFS (GFS or HDFS) in a MapReduce system?
- 3) What is the difference between a NameNode and a DataNode in HDFS?
- 4) What is a data block in HDFS and how is it replicated across data nodes?