

TDT4305 2021 - Assignment 2

Spark

- 1) What are some of the differences between Spark and MapReduce?
- 2) What are *narrow* and *wide dependencies*, and how do they relate to a *stage*?
- 3) What are the three ways of storing persistent RDDs and how are they each useful?
- 4) What is checkpointing and when is it typically used?
- 5) What is the purpose of lazy evaluation?