TDT4305 2021 - Assignment 2 solutions

Spark

1) What are some of the differences between Spark and MapReduce?

   ○ *See [Zaharia] section 1.*

2) What are *narrow* and *wide dependencies*, and how do they relate to a *stage*?

   ○ *See [Zaharia] section 5.1.*

3) What are the three ways of storing persistent RDDs and how are they each useful?

   ○ *See [Zaharia] section 5.3.*

4) What is checkpointing and when is it typically used?

   ○ *See [Zaharia] section 5.4.*

5) What is the purpose of lazy evaluation?

   ○ *See [LearningSpark] page 29.*