TDT4225 – Exercise 1 (theory)

Deadline: 17. Sep. Deliver individually on Blackboard. Please deliver in pdf format.

1. **SSD:** How does the Flash Translation Level (FTL) work in SSDs?
2. **SSD:** Why are sequential writes important for performance on SSDs?
3. **SSD:** Discuss the effect of alignment of blocks to SSD pages.
4. **RocksDB**: Describe the layout of MemTable and SSTable of RocksDB.
5. **RocksDB:** What happens during compaction in RocksDB?
6. **LSM-**trees vs B+-trees. Give some reasons for why LSM-trees are regarded as more efficient than B+-trees for large volumes of inserts.
7. Regarding fault tolerance, give a description of what hardware, software and human errors may be?
8. Compare SQL and the document model. Give advantages and disadvantages of each approach. Give an example which shows the problem with many-to-many relationships in the document model, e.g., how would you model that a paper has many sections and words, and additionally it has many authors, and that each author with name and address has written many papers?
9. When should you use a graph model instead of a document model? Explain why.
   Give an example of a typical problem that would benefit from using a graph model.
10. **Column compression**: You have the following values for a column:
    43 43 43 87 87 63 63 32 33 33 33 33 89 89 89 33
    a) Create a bitmap for the values.
    b) Create a runlength encoding for the values

10. We have different binary formats / techniques for sending data across the network:

    - MessagePack
    - Apache Thrift
    - Protocol Buffers
    - Avro

    In case we need to do schema evolution, e.g., we add a new attribute to a Person structure: Labour union, which is a String. How is this supported by the different systems? How is forward and backward compatibility supported?