

i **Framside / Front page**

Department of Computer Science

Examination paper for TDT4305 Big Data Architecture

Examination date: 23th of May 2020

Examination time (from-to): 09:00 – 13:00

Permitted examination support material: All support material is allowed

Academic contact during examination:

Phone: 99 02 76 56

Technical support during examination: [Orakel support services](#)

Phone: 73 59 16 00

OTHER INFORMATION

If a question is unclear/vague – make your own assumptions and specify in your answer the premises you have made. Only contact academic contact in case of errors or insufficiencies in the question set.

Saving: Answers written in Inspira are automatically saved every 15 seconds. If you are working in another program remember to save your answer regularly.

Cheating/Plagiarism: The exam is an individual, independent work. Examination aids are permitted. All submitted answers will be subject to plagiarism control. [Read more about cheating and plagiarism here.](#)

Citations: All questions are expected to be answered "in own words" in order to show understanding.

Notifications: If there is a need to send a message to the candidates during the exam (e.g. if there is an error in the question set), this will be done by sending a notification in Inspira. A dialogue box will appear. You can re-read the notification by clicking the bell icon in the top right-hand corner of the screen. All candidates will also receive an SMS to ensure that nobody misses out on important information. Please keep your phone available during the exam.

Weighting: As shown for each problem. All sub-problems within a problem have equal weight.

ABOUT SUBMISSION

Your answer will be submitted automatically when the examination time expires and the test closes, if you have answered at least one question. This will happen even if you do not click “Submit and return to dashboard” on the last page of the question set. You can reopen and edit your answer as long as the test is open. If no questions are answered by the time the examination time expires, your answer will not be submitted.

Withdrawing from the exam: If you wish to submit a blank test/withdraw from the exam, go to the menu in the top right-hand corner and click “Submit blank”. This can not be undone, even if the test is still open.

Accessing your answer post-submission: You will find your answer in Archive when the examination time has expired.

Problem 1 – Systems for data streams (15%)

- | | Guarantee | Example |
|-----------|-----------|---------|
| AsterixDB | | |
| Spark | | |
| Storm | | |

-

Words: 0

2/7

Problem 2 – Handling of data stream (15%)

1. Explain how you would solve this problem using standing queries.
2. Could this problem be solved using bit counting? Justify your answer.
3. Assume you will use buckets to solve this problem. How does the concept of buckets work? How would you solve this problem using the principle of buckets?

3/7

4 Oppgave 4

Problem 4 – Hadoop and Spark – 15 %

1. Explain how writing to a file is performed in HDFS.
2. Explain the concepts "narrow dependency" and "wide dependency", and how these affects performance.

Fill in your answer here

Maximum marks: 15

Problem 5 – MinHashing – 20 %

Fill in your answer here

Maximum marks: 20

Problem 6 – Adwords – 15 %

Fill in your answer here

Maximum marks: 15