

## i Framside / Front page

Institutt for datateknologi og informatikk

Eksamensoppgave i TDT4305 Big Data Architecture

Eksamensdato: 23. mai 2020

Eksamenstid (fra-til): 09:00 – 13:00

Hjelpemiddelkode/Tillatte hjelpemidler: A / Alle hjelpemidler tillatt

Faglig kontakt under eksamen:

Tlf.: 99 02 76 56

Teknisk hjelp under eksamen: [NTNU Orakel](#)

Tlf: 73 59 16 00

### ANNEN INFORMASJON:

Gjør dine egne antagelser og presiser i besvarelsen hvilke forutsetninger du har lagt til grunn i tolkning/avgrensing av oppgaven. Faglig kontaktperson skal kun kontaktes dersom det er direkte feil eller mangler i oppgavesettet.

**Lagring:** Besvarelsen din i Inspira Assessment lagres automatisk. Jobber du i andre programmer – husk å lagre underveis.

**Juks/plagiat:** Eksamen skal være et individuelt, selvstendig arbeid. Det er tillatt å bruke hjelpemidler. Alle besvarelser blir kontrollert for plagiat. [Du kan lese mer om juks og plagiering på eksamen her.](#)

**Kildehenvisninger:** Alle oppgaver skal besvares "med egne ord" for å vise forståelse.

**Varslinger:** Hvis det oppstår behov for å gi beskjeder til kandidatene underveis i eksamen (f.eks. ved feil i oppgavesettet), vil dette bli gjort via varslinger i Inspira. Et varsel vil dukke opp som en dialogboks på skjermen i Inspira. Du kan finne igjen varselet ved å klikke på bjella øverst i høyre hjørne på skjermen. Det vil i tillegg bli sendt SMS til alle kandidater for å sikre at ingen går glipp av viktig informasjon. Ha mobiltelefonen din tilgjengelig.

**Vekting av oppgavene:** Som vist i oppgavesettet. Alle deloppgaver innenfor en oppgave teller likt.

### OM LEVERING:

**Besvarelsen din leveres automatisk når eksamenstida er ute og prøven stenger**, forutsatt at minst én oppgave er besvart. Dette skjer selv om du ikke har klikket «Lever og gå tilbake til Dashboard» på siste side i oppgavesettet. Du kan gjenåpne og redigere besvarelsen din så lenge prøven er åpen. Dersom ingen oppgaver er besvart ved prøveslutt, blir ikke besvarelsen din levert.

**Trekk fra eksamen:** Ønsker du å levere blankt/trekke deg, gå til hamburgermenyen i øvre høyre hjørne og velg «Lever blankt». Dette kan ikke angres selv om prøven fremdeles er åpen.

**Tilgang til besvarelse:** Du finner besvarelsen din i Arkiv etter at sluttida for eksamen er passert

### Oppgave 1 – Systemer for datastrømmer (15%)

- |           | Garanti | Eksempel |
|-----------|---------|----------|
| AsterixDB |         |          |
| Spark     |         |          |
| Storm     |         |          |

- 

Words: 0

2/7

## 3

$$sim(x, y) = \frac{\sum_{s \in S_{xy}} (r_{xs} - \bar{r}_x)(r_{ys} - \bar{r}_y)}{\sqrt{\sum_{s \in S_{xy}} (r_{xs} - \bar{r}_x)^2} \sqrt{\sum_{s \in S_{xy}} (r_{ys} - \bar{r}_y)^2}}$$

**Skriv ditt svar her...**

4/7

#### 4 Oppgave 4

### Oppgave 4 – Hadoop og Spark – 15 %

1. Forklar hvordan skriving til fil foregår i HDFS.
2. Forklar konseptene "narrow dependency" og "wide dependency", og hvordan disse påvirker ytelse.

**Skriv ditt svar her...**

Maks poeng: 15

### Oppgave 5 – MinHashing – 20 %

**Skriv ditt svar her...**

Maks poeng: 20

## 6 Oppgave 6

### Oppgave 6 – Adwords – 15 %

Forklar "Balance-"algoritmen (i kontekst av Adwords).

**Skriv ditt svar her...**

Maks poeng: 15