

**Institutt for datateknologi og informatikk****Eksamensoppgave i TDT4150 Avanserte databasesystemer****Eksamensdato:** 2. juni 2021**Eksamenstid (fra-til):** 09:00-13:00**Hjelpemiddelkode/Tillatte hjelpemidler:** A / Alle hjelpemidler tillatt**Faglig kontakt under eksamen:** Norvald H. Ryeng**Tlf:** 971 74 980**Teknisk hjelp under eksamen:** NTNU Orakel**Tlf:** 73 59 16 00

Får du tekniske problemer underveis i eksamen, må du ta kontakt for teknisk hjelp snarest mulig, og senest innen eksamenstida løper ut. Kommer du ikke gjennom umiddelbart, hold linja til du får svar.

**ANNEN INFORMASJON**

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

**Juks/plagiat:** Eksamen skal være et individuelt, selvstendig arbeid. Det er tillatt å bruke hjelpemidler, men vær obs på at du må følge eventuelle anvisningen om kildehenvisninger under. Under eksamen er det ikke tillatt å kommunisere med andre personer om oppgaven eller å distribuere utkast til svar. Slik kommunikasjon er å anse som juks.

Alle besvarelser blir kontrollert for plagiat. [Du kan lese mer om juks og plagiering på eksamen her.](#)

**Kildehenvisninger:** Husk å ta med kildereferanse dersom du siterer fra kilder i besvarelsen.

**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:** Alle oppgaver teller like mye.

## OM LEVERING

**Slik svarer du på oppgavene:** Alle oppgaver som *ikke* er av typen filopplasting, skal besvares direkte i Inspera. I Inspera lagres svarene dine automatisk hvert 15. sekund.

NB! Klipp og lim fra andre programmer frarådes, da dette kan medføre at formatering og elementer (bilder, tabeller etc.) vil kunne gå tapt.

**Automatisk innlevering:** 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. Dette vil anses som “ikke møtt” til eksamen.

**Trekk/avbrutt eksamen:** Blir du syk under eksamen, eller av andre grunner ønsker å levere blankt/avbryte eksamen, 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.

## 1 Problem 1

PostgreSQL og MySQL har forskjellige prosessmodeller. PostgreSQL bruker en OS-prosess per sesjon, mens MySQL bruker en OS-tråd per sesjon.

Forklar fordeler og ulemper ved de to modellene.

**Skriv ditt svar her**

## 2 Problem 2

Både ved prosess per sesjon og tråd per sesjon er det mulig å bruke pooling (process pooling eller thread pooling). Hva er fordelene ved pooling?

**Skriv ditt svar her**

### 3 Problem 3

Et DBMS med de samme egenskapene som System R mottar denne spørringen:

```
SELECT *  
FROM t1, t2, t3  
WHERE t1.a = t2.a AND t2.b = t3.b  
ORDER BY t1.c + 2 * t2.c + 0.5 * t3.c  
LIMIT 10;
```

List alle interessante rekkefølger (interesting orders) og spesifiser for hver rekkefølge hvorfor den er interessant.

**Skriv ditt svar her**

### 4 Problem 4

Systemet i problem 3 blir utvidet med støtte for rank join. Hvilke nye rekkefølger i spørringen blir interessante? Forklar hvorfor.

**Skriv ditt svar her**

### 5 Problem 5

Forklar problemet med tomme join-resultat (empty join results) med tradisjonell sampling-basert kardinalitetsestimering. Hva er det og hvorfor oppstår det?

**Skriv ditt svar her**

### 6 Problem 6

Hva gjør indeks-basert join-sampling (index based join sampling) for å løse problemet med tomme join-resultat (empty join results)?

**Skriv ditt svar her**

## 7 Problem 7

Hva er pipelining i forbindelse med utføring av spørringer? Gi et eksempel på en ikke-pipelining operasjon og forklar hvorfor den ikke er pipelining.

**Skriv ditt svar her**

## 8 Problem 8

Hvordan påvirker støtte for interessante rekkefølger (interesting order tracking) bruken av pipelining operasjoner?

**Skriv ditt svar her**

## 9 Problem 9

Forklar hvordan slettenoder (delete node entries) i LSM-trær fungerer og hvordan de påvirker oppslag (lookups).

**Skriv ditt svar her**

## 10 Problem 10

Hva er formålet med join-indekser (join indexes) i C-Store? Hvorfor er join-indekser ikke nødvendige i rad-basert lagring (row store)?

**Skriv ditt svar her**

## 11 Problem 11

Anta et distribuert DBMS spesiallaget for å lagre kalenderhendelser (møter, avtaler, osv.). En bruker kan ha flere telefoner, nettbrett og datamaskiner. Alle enhetene har en komplett kopi av brukerens kalender og betraktes som noder i det samme DBMS-et. I tillegg er det en node som kjører som en skytjeneste som også har en komplett kopi av dataene. Selv om en enhet er frakoblet kan brukeren se og redigere kalenderhendelser på den enheten. Dataene blir synkronisert med de andre nodene som er påkoblet når enheten kobles til og får kontakt med skytjenesten, som tar rollen som koordinator for synkroniseringen.

Hvor vil du plassere dette systemet langs aksene for transaksjonslokasjon (transaction location/where) og synkroniseringspunkt (synchronization point/when). Forklar hvorfor.

**Skriv ditt svar her**

## 12 Problem 12

Gi et eksempel på hvordan inkonsistens kan oppstå i systemet beskrevet i problem 11 når enheter opererer mens de er frakoblet.

**Skriv ditt svar her**

## 13 Problem 13

Hvlike av de tre egenskapene i CAP-teoremet har systemet i problem 11? Forklar hvorfor.

**Skriv ditt svar her**

## 14 Problem 14

Et DBMS mottar denne spørringen om hoteller som skal vises på et kart på skjermen på en mobiltelefon. Kartutsnittet som er synlig på skjermen er boksen som kalles @map\_viewport.

```
SELECT *  
FROM hotel JOIN country ON hotel.country_id = country.id  
WHERE  
    country.population > 1000000  
    AND  
    ST_Within(hotel.position, @map_viewport)  
    AND  
    hotel.star_rating = 4  
ORDER BY hotel.name;
```

Det boolske uttrykket ST\_Within(a, b) er sant dersom a er i b.

Anta at det er indekser på alle kolonner nevnt i spørringen og at indeksene støtter vindusspørringer (window queries), punktspørringer (point queries), avstandsskanning (distance scan), oppslag på likhet (equality lookup), oppslag på område (index range scan) og indeksskanning (full index scan).

List alle sargbare predikater (sargable predicates) i spørringen og oppgi for hvert predikat hvilken type indeksaksess som vil bli brukt.

**Skriv ditt svar her**