

Institutt for datateknikk og informasjonsvitenskap

## **Eksamensoppgave i TDT4225 Lagring og behandling av store datamengder**

**Faglig kontakt under eksamen: Kjell Bratbergsengen**

**Tlf.: 90617 185 / 7359 3439**

**Eksamensdato: Lørdag 18. mai 2013**

**Eksamenstid (fra-til): 0900 - 1300**

**Hjelpemiddelkode/Tillatte hjelpemidler: D**

*Ingen trykte eller håndskrevne hjelpemiddel er tillatt.*

*Bestemt, enkel kalkulator tillatt*

**Annen informasjon:**

*Sensur innen: tirsdag 11. juni 2013.*

**Målform/språk: Bokmål**

**Antall sider: 3**

**Antall sider vedlegg: 0**

**Kontrollert av:**

---

Dato

Sign

# TDT4225 Lagring og behandling av store datamengder

Lørdag 18. mai 2013, kl. 0900-1300

## Oppgave 1, Sekvensoptimalisering av I/O (15 %)

For å effektivisere I/O kan diskkontroller bruke forskjellig strategi for å betjene I/O-ordrene.

a) Forklar strategiene FCFS-first come first served, SSTF-shortest seek time first og SCAN.

b) Følgende operasjoner skal utføres på en disk med 1000 sylindre, bare sylindre-delen av adressen er gitt. Lesehodets starposisjon er sylindre 500. Hva blir total posisjoneringsdistanse for de forskjellige strategiene beskrevet i oppgave a) når adressene er:

903, 396, 442, 112, 222, 741, 832.

Anta at SSTF alltid har tre forespørslers i køen (unntatt for de siste to operasjonene). SCAN samler opp alle forespørslene før den starter behandlingen.

## Oppgave 2, Videoserver (20 %)

En videoserver skal levere film til individuelle brukere som ser på forskjellige filmer. Hver film krever en tilnærmet konstant datastrøm over tid på 3 MB/s. Lokalt har hver skjerm (bruker) et lokalt lager som holder data for 15 sekunder film. Filmene ligger på en SSD (solid state disk). Disken er koplet til en systembuss med en optisk kabel som kan overføre opp til 300 MB/s. Systembussen har en kapasitet på 800 MB/s. Anta at maskinens arbeidslager kan "holde følge" med systembussens kapasitet. Arbeidslager er 1 GB. Dataoverføringer styres av en DMA-enhet.

a) Vis hvordan data flyter over de forskjellige systemdelene.

b) SSD har et tidsforbruk ved lesing som følger formelen  $t(x)=a+bx$ .  $x$  er blokkstørrelse i antall byte.  $a$  er 0,1 millisekund og  $b$  regnes ut på bakgrunn av at lesehastigheten er 200 MB/s. Hvilken systemdel (disk, optisk kabel, systembuss) er flaskehals når ingen av disse skal utnyttes til mer enn 70 % av toppkapasitet.

c) Hvor mange samtidige skjermer (brukere) kan serveren betjene? Hvilken blokkstørrelse har du valgt å bruke i overføringene.

## Oppgave 3, LRU og Clock (10 %)

a) Forklar LRU-algoritmen ? Hvorfor er algoritmen dårlig ved lesing av en sekvensiell fil?

b) Forklar Clock-algoritmen.

c) Vi har et arbeidslager med plass for 4 bufre. Gitt følgende referansestreng for objekter (blokker):

a, b, c, d, b, d, e, a, b, b, d, c, e

i) Hvor mange blokker må leses når LRU-algoritmen brukes.

ii) Hvor mange blokker må leses når Clock-algoritmen brukes.

Vis hvordan du kommer fram til resultatet.

#### Oppgave 4, Lagring av rasterbilder (20 %)

Et stort bilde er lagret i blokker, hver blokk inneholder et bildekvadrat. Hvert piksel i bildet tar 4 byte.

- Hvis bildet lagres i blokker som hver tar 32 KB (32768 byte) hvor stor blir da sidekanten i bildekvadratet som en blokk inneholder?
- Fra bildet tas det utsnitt, vilkårlig plassert, men med sidekanter som er parallelle med bildekantene. Størrelsen på utsnittet er 2000 x 2000 piksler. Hvor stort blir I/O-volumet for uthenting av slike utsnitt i gjennomsnitt? Blokkstørrelsen er fortsatt 32 KB.
- Bildet lagres på disk som roterer 15000 ganger per minutt. Hvert spor lagrer 0,5 MB. Hva blir overføringstiden for å hente et utsnitt som beskrevet i punkt b)?

#### Oppgave 5, UNIX (15 %)

- Forklar datastrukturen som brukes av UNIX filsystem for å komme fra logisk intern adresse innen en fil til en fysisk adresse på disk.
- Et program skriver en string på 1000 byte til fil på følgende adresser:  
0, 5000, 10000, 15000, 20000.  
Hvor mange blokker blir allokeret til filen. Installasjonen bruker en blokkstørrelse på 4 KB (4096 byte).

#### Oppgave 6, Relasjonsalgebra (20 %)

	Resultat	Basetabeller			Mellomresultat		
	R	A	B	C	$A*B=B*A$	$A*C=C*A$	$B*C=C*B$
Nøkkellengde i byte	8	8	8	8	8	8	8
Postlengde i byte	200	1200	400	700	200	200	200
Antall poster	300 000	500 000	1 200 000	50 000	400 000	200 000	100 000
Tabellvolum i MB	60	600	480	35	80	40	20

En skal utføre operasjonen  $R=A*B*C$ . Informasjon i tabellen over. Nøkkellengde gjelder join-operasjonens nøkkel.

- Finne den sekvens av operasjoner som gir minst I/O-volum. Anta at mellomresultater er lagret på fil. Ikke vurder join-algoritmer eller ta hensyn til eventuelle interne mellomlagringer i operasjonene i denne deloppgaven.
- Finne den beste algoritmen for å gjøre operasjonen  $U=A*B$ . Resultatpostene i  $U$  henter 100 byte fra hver av operandene. Tilgjengelig arbeidslager  $WS$  er 10 MB.