

### 3. VJEŽBA

#### Uvodne napomene:

- upiti (SQL naredbe) i podaci koji će se koristiti u ovoj vježbi ne moraju biti logički ispravni i ne moraju zadovoljavati integritetska ograničenja
- sintaksu naredbi koje se koriste u ovim vježbama nije potrebno učiti napamet, ali se očekuje razumijevanje njihovog značenja i njihovih rezultata
- baza podataka labprof3 sadrži dovoljno prikladnih relacija i podataka za provođenje eksperimenata u ovim vježbama. Ne preporuča se, ali dopušteno je, kreirati i podacima puniti i vlastite relacije. Ako kreirate vlastite relacije, nemojte definirati primarne, alternativne i strane ključeve, kako bi se izbjeglo kreiranje neželjenih indeksa koji bi mogli utjecati na rezultate eksperimenata. U takvom slučaju rješenjima treba priložiti sve što je potrebno da bi se takve dodatne relacije napunile podacima koje ste koristili u eksperimentima.
- SQL Server statističke podatke po definiciji generira automatski, izvan kontrole korisnike. To nije prikladno tijekom provođenja eksperimenata u ovoj vježbi, stoga je automatsko generiranje statističkih podataka za bazu podataka labprof3 isključeno već tijekom kreiranja baze (pogledati naredbe ALTER DATABASE u labprof3.sql). Statistički podaci će se, bez obzira na to, i dalje generirati automatski kada se kreira neki indeks jer to nije moguće isključiti.
- za SQL Server je karakteristično:
  - za razliku od nekih drugih sustava, parametri na temelju kojih se provodi optimizacija upita i odabir fizičkih operatora su vrlo slabo dokumentirani i relativno se često mijenjaju, ponekad od verzije do verzije sustava
  - korisnik nema mogućnost odabira kategorije statističkih podataka koje će se voditi za pojedine objekte u bazi podataka

#### Korisne informacije:

- Isključiti ili uključiti automatsko kreiranje i ažuriranje statističkih podataka za bazu podataka *dbname*.

```
ALTER DATABASE dbname SET AUTO_CREATE_STATISTICS OFF/ON;
```

```
ALTER DATABASE dbname SET AUTO_UPDATE_STATISTICS OFF/ON;
```

Postavljanjem ovih svojstava na OFF, statistički podaci se prestaju automatski generirati i ažurirati, osim za indekse te za relacije i atribute nad kojima je kreiran indeks. Statistički podaci, odnosno posebni objekti u kojima se bilježe statistički podaci, generirani u trenutku kreiranja indeksa, mogu se uništiti jedino uništavanjem tog indeksa.

- Za koje relacije/atribute trenutačno postoje posebni objekti u kojima se bilježe statistički podaci može se provjeriti upitom

```
SELECT stats.name, auto_created, user_created, tables.name, columns.name
FROM sys.stats
JOIN sys.tables ON stats.object_id = tables.object_id
JOIN sys.stats_columns ON stats_columns.object_id = stats.object_id
AND stats_columns.stats_id = stats.stats_id
JOIN sys.columns ON columns.object_id = stats_columns.object_id
AND columns.column_id = stats_columns.column_id;
```

## 1. zadatak: Kreirati testnu bazu podataka labprof3

Procjena potrebnog vremena: 5 minuta

U rješenjima je potrebno dostaviti: ništa

Preuzeti datoteke \*.csv i izvršiti skripta labprof3.sql.

Uočiti da skripta sadrže naredbe koje isključuju automatsko kreiranje i ažuriranje statističkih podataka namijenjenih optimizatoru (ALTER DATABASE labprof3 SET ...). Ovdje je to napravljeno zato da se u zadacima i eksperimentima može kontrolirati koje statističke podatke optimizator koristi pri odabiru fizičkih operatora.

Tijekom nadziranih provjera/vježbi očekuju se sljedeća znanja i vještine:

- objasniti na koji način i zašto je ovdje isključeno automatsko kreiranje i ažuriranje statističkih podataka

## 2. zadatak: Statistički podaci (koji se uvijek vode) za relaciju i indeks

Procjena potrebnog vremena: 15 minuta

U rješenjima je potrebno dostaviti: kratke odgovore na pitanja u zadacima

Izvršiti jednu po jednu naredbu i ukratko odgovoriti na pitanja.

**2.1.** Koji se statistički podaci za koju relaciju dobiju izvršavanjem sljedećeg upita? Kojim su simbolima vrijednosti `page_count` i `record_count` označene u predavanjima iz predmeta SBP?

```
SELECT object_name(object_id) AS table_name, page_count, record_count
FROM sys.dm_db_index_physical_stats (DB_ID('labprof3'), OBJECT_ID('salesOrderItem'), 0, NULL, 'DETAILED');
```

**2.2.** Kreirati indeks:

```
CREATE INDEX i1 ON salesOrderItem (orderQty);
```

Koji se statistički podaci za koji indeks dobiju izvršavanjem sljedećeg upita? Kojim simbolom je vrijednost `index_depth` označena u predavanjima iz predmeta SBP?

```
SELECT name AS index_name, index_type_desc, index_depth, index_level, page_count, record_count
FROM sys.dm_db_index_physical_stats (
    DB_ID('labprof3')
    , OBJECT_ID('salesOrderItem')
    , NULL
    , NULL
    , 'DETAILED') AS t1
JOIN sys.indexes AS t2
    ON t1.index_id = t2.index_id
    AND t1.object_id = t2.object_id
    AND name IS NOT NULL;
```

**2.3.** Uništiti indeks kreiran u točki 2.2.

Tijekom nadziranih provjera/vježbi očekuju se sljedeća znanja i vještine:

- objasniti značenje statističkih podataka dobivenih izvršavanjem ovdje prikazanih upita

### 3. zadatak: Kreiranje i pregled dodatnih statističkih podataka

Procjena potrebnog vremena: 30 minuta

U rješenjima je potrebno dostaviti: kratke odgovore na pitanja u zadacima

Osim statističkih podataka koji se uvijek vode uz relaciju ili indeks, moguće je kreirati dodatne objekte koji sadrže statističke podatke. Ti dodatni objekti se kreiraju automatski, osim onda kada se isključi automatsko kreiranje statističkih objekata. Tada se objekti koji sadrže statističke podatke mogu kreirati posebnim naredbama ili se kreiraju automatski u trenutku kada se kreira neki indeks.

Izvršiti jednu po jednu naredbu i odgovoriti na pitanja.

#### 3.1. Čemu služi sljedeća naredba? Što znači WITH FULLSCAN?

```
CREATE STATISTICS stat1 ON salesOrderItem (orderQty) WITH FULLSCAN;
```

#### 3.2. Čemu služi sljedeća naredba? Što znači WITH ALL/COLUMNS/INDEX?

```
UPDATE STATISTICS salesOrderItem WITH ALL;
```

#### 3.3. Što predstavljaju veličine Name, Rows, Rows Sampled, Steps u rezultatu sljedeće naredbe?

```
DBCC SHOW_STATISTICS (salesOrderItem, stat1) WITH STAT_HEADER;
```

#### 3.4. Što predstavljaju veličine All density, columns u rezultatu sljedeće naredbe?

```
DBCC SHOW_STATISTICS (salesOrderItem, stat1) WITH DENSITY_VECTOR;
```

#### 3.5. Što predstavljaju veličine RANGE\_HI\_KEY, RANGE\_ROWS, EQ\_ROWS, DISTINCT\_RANGE\_ROWS, AVG\_RANGE\_ROWS u rezultatu sljedeće naredbe?

```
DBCC SHOW_STATISTICS (salesOrderItem, stat1) WITH HISTOGRAM;
```

#### 3.6. Kojom naredbom se može uništiti objekt sa statističkim podacima koji je kreiran u točki 3.1? **Uništite taj objekt.**

#### 3.7. Kreirati sljedeći indeks i odgovoriti koje je ime objekta sa statističkim podacima koji se automatski kreirao kao posljedica kreiranja tog indeksa? Sadrži li taj objekt sve podatke kao i objekt kreiran u točki 3.1?

```
CREATE INDEX i1 ON salesOrderItem (orderQty);
```

#### 3.8. Kojom naredbom se može uništiti objekt sa statističkim podacima koji je kreiran istovremeno kad je kreiran indeks u točki 3.7? **Uništite taj objekt.**

Tijekom nadziranih provjera/vježbi očekuju se sljedeća znanja i vještine:

- objasniti što znače ovdje testirane naredbe i rezultati izvršavanja tih naredbi (samo oni rezultati koji su istaknuti u zadacima)
- na zahtjev kreirati, prikazati i uništiti objekt sa statističkim podacima za neku relaciju/atribut ili indeks

## 4. zadatak: Plan izvršavanja

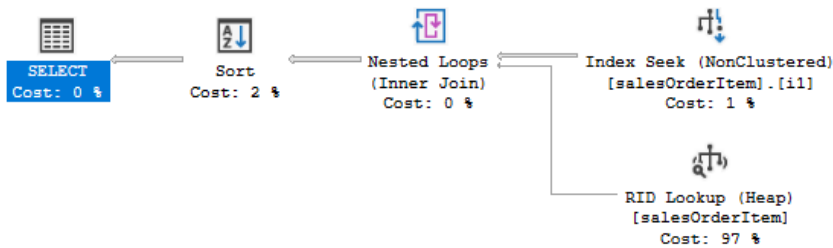
Procjena potrebnog vremena: 20 minuta

U rješenjima je potrebno dostaviti: ništa

Kreirati indeks: `CREATE INDEX i1 ON salesOrderItem (orderQty)`

U alatu SSMS prikazati plan izvršavanja za sljedeći upit:

```
SELECT * FROM salesOrderItem WHERE orderQty > 19 ORDER BY salesOrderID;
```



U detaljima svake od operacije, npr.

Index Seek (NonClustered)	
Scan a particular range of rows from a nonclustered index.	
Physical Operation	Index Seek
Logical Operation	Index Seek
Estimated Execution Mode	Row
Storage	RowStore
Estimated Operator Cost	0,0035163 (1%)
Estimated I/O Cost	0,003125
Estimated Subtree Cost	0,0035163
Estimated CPU Cost	0,0003913
Estimated Number of Executions	1
Estimated Number of Rows	213
Estimated Number of Rows to be Read	213
Estimated Row Size	19 B
Ordered	True
Node ID	3
<b>Object</b> [labprof3].[dbo].[salesOrderItem].[i1]	
<b>Output List</b> Bmk1000: [labprof3].[dbo].[salesOrderItem].orderQty	
<b>Seek Predicates</b> Seek Keys[1]: Start: [labprof3].[dbo].[salesOrderItem].orderQty > Scalar Operator((19))	

uočiti sljedeće vrijednosti:

- Physical Operation
- Logical Operation
- Estimated Operator Cost
- Estimated Subtree Cost
- Estimated I/O Cost
- Estimated Number of Rows

Tijekom nadziranih provjera/vježbi očekuju se sljedeća znanja i vještine:

- prikazati i komentirati plan izvršavanja za jednostavnu SQL naredbu
- objasniti vrijednosti Physical Operation, Logical Operation, Estimated Operator Cost, Estimated Subtree Cost, Estimated I/O Cost, Estimated Number of Rows

## 5. zadatak: Fizički operatori

Procjena potrebnog vremena: 180 minuta

U rješenjima je za zadatke 5.2 - 5.11 potrebno dostaviti: **tekst zadatka**, pripadne SQL naredbe, **čitljivu** sliku plana izvršavanja i kratko **obrazloženje** zašto je optimizator odabrao baš taj fizički operator. Upit treba napisati tako da sustav samostalno odabere u zadatku traženi operator, a ne da ga na korištenje operatora navede eksplicitna uputa optimizatoru, tzv. *query hint*. Primjer rješenja jednog zadatka (5.1) prikazan je na kraju poglavlja.

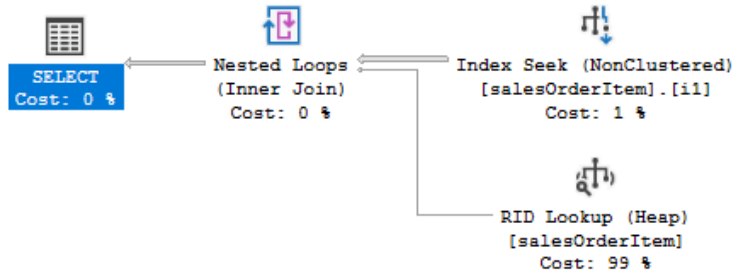
- 5.1. Napisati upit koji koristi operator *index-seek* + *RID lookup* (u predavanjima se naziva *index-scan*) - za ovaj zadatak ne treba dostavljati rješenje - pogledati rješenje na kraju ovog poglavlja.
- 5.2. Napisati upit koji koristi SQL Server operator *index-seek* bez *RID lookup* (u predavanjima se naziva *key-only index scan*).
- 5.3. Napisati upit koji koristi SQL Server operator *index-scan* (u predavanjima se naziva *key-only index scan*).
- 5.4. Proučiti što su to indeksi s dodanim atributima (*Indexes with Included Columns*). Indeks koji se koristio u rješenju zadatka 5.1, u ovom zadatku treba prepraviti u indeks s dodanim atributima, na način koji će osigurati izvršavanje istog upita korištenjem samo operatora *index-seek* (dakle bez dodatnog *RID Lookup*). Usporedite veličinu memorije koju koriste indeksi te ukupni procijenjeni trošak izvršavanja upita u zadacima 5.1. i 5.4.
- 5.5. Napisati upit koji koristi operator *table-scan*.
- 5.6. Napisati upit koji koristi SQL Server operator *index-seek* (*clustered*).
- 5.7. Napisati upit koji koristi SQL Server operator *index-scan* (*clustered*).
- 5.8. Napisati upit koji *non-clustered* indeks koristi za operator *index-seek* za dohvrat n-torki koje se nalaze u listovima drugog, *clustered* indeksa.
- 5.9. Napisati upit koji sortiranje dohvaćenih podataka obavlja koristeći SQL Server operator *index-scan* (u predavanjima se naziva *key-only index scan*).
- 5.10. Napisati upit koji koristi SQL Server operator *index-seek* + *RID lookup*, pri čemu se isti indeks koristi i za sortiranje.
- 5.11. Napisati upit koji koristi SQL Server operator *index-seek* + *RID lookup*, pri čemu se isti indeks neće moći koristiti i za sortiranje.

**- primjer rješenja zadatka -**

**5.1.** Napisati upit koji koristi operator *index-seek* + *RID lookup* (u predavanjima se naziva *index-scan*) - za ovaj zadatak ne treba dostavljati rješenje - pogledati rješenje na kraju ovog poglavlja.

```
CREATE INDEX i1 ON salesOrderItem (orderQty);  
UPDATE STATISTICS salesOrderItem WITH ALL;  
SELECT salesOrderID, orderQty FROM salesOrderItem WHERE orderQty > 19;  
-- vratiti na staro stanje  
DROP INDEX i1 ON salesOrderItem;
```

Plan izvršavanja:



Obrazloženje: Za predikat selekcije postoji odgovarajući indeks. Pomoću B-stabla pronalaze se (Index Seek) odgovarajuće kazaljke na n-torke u podacima s blokovima, koje se moraju dohvatiti (RID Lookup) iz blokova s podacima jer ključevi u B-stablu ne sadrže sve podatke koji su navedeni u SELECT listi. U rezultatu se očekuje relativno mali broj n-torki, stoga je korištenje indeksa opravdano.

**- kraj primjera rješenja zadatka -**

Tijekom nadziranih provjera/vježbi očekuju se sljedeća znanja i vještine:

- u nekoj zadanoj bazi podataka, napisati upit koji će dovesti do korištenja zadanog fizičkog operatora, obrazložiti princip funkcioniranja tog operatora, razlog zašto je optimizator odabrao taj operator, prikazati i komentirati plan izvršavanja upita.