

SQL Server Import, SQL Server Profiler, Data analyse og anvendelse af INDEX

Denne opgave handler om at optimere løsninger på baggrund af dataanalyse med brug af INDEX.

Først skal der opbygges en database med en masse data. Databasen kan hedde *Performance* og indeholder en tabel *Random* med to felter *Id* og *RandomNumber*.

Dernæst skal der genereres data til denne tabel. Der skal i alt være 1 million records med to heltal, hvor det første heltal er identifier – nøglen – og andet tal er et tilfældigt tal mellem 0 og 9.999. Med et C#-program oprettes en tekstfil vha. biblioteksklassen *File*. Til denne fil skrives 1 million linjer med de to genererede heltal med et komma imellem.

Nu skal disse data importeres til tabellen *Random*. Det kan enten gøres med en *BULK INSERT* SQL-sætning eller *SQL Server Import and Export Wizard*. Det er ikke lykkedes mig at få SQL-sætningen til at fungere sammen med tekstfilens syntaks¹, så umiddelbart anbefaler jeg Wizard'en², men det vil være skønt, hvis det også kan lade sig gøre med BULK-sætningen.

Hermed er tabellen fyldt med 2 millioner tal, og så længe man kun skal slå op ud fra nøglen, går det hurtigt, hvorimod analyse af de tilfældige tal kan tage tid. Det er det, der nu skal undersøges nærmere.

For at få anvendelserne af databasen *Performance* målt løbende, skal der startes et andet program og dermed en anden proces, som afvikles sideløbende med SSMS. Det er *SQL Server Profiler*. Programmet startes fra SSMS under menupunktet *Tools*. Her kan man se alt, hvad der foregår i Database Engine, men nu er der udelukkende anvendelsen af *Performance*, der har interesse.

Nu skal der laves følgende forespørgsler til databasen (husk at gemme løsningerne til senere brug):

- Søg efter et bestemt tilfældigt tal, fx 4711
- Lav en oversigt over alle de tilfældige tal og hvor mange gange de hver især forekommer, sorteret efter de tilfældige tal (dette kan også benyttes som et VIEW til løsning af de næste forespørgsler)
- Find hvor mange gange det eller de sjældneste tilfældige tal forekommer
- Find hvor mange gange det eller de hyppigste tilfældige tal forekommer
- Benyt de sidste to resultater til at finde hvilke tal det er, der er sjældnest eller hyppigst

Alt er relativt, men af tallene (CPU, Reads, Duration) kan man se, at der bruges nogle ressourcer. Tilføj nu et INDEX til tabellen *Random*. Der findes to typer index: *clustered* og *nonclustered*. I min udgave af SSMS er det umiddelbart kun muligt at benytte *nonclustered* via Object Explorer, så afprøv det.³ Det er feltet *RandomNumber*, der laves index på.

Afprøv nu de samme forespørgsler som tidligere og dokumenter forskellene.

PS: Man kan også lave et INDEX på view'et, men det er kun på 10.000 records. Og hvis jeres maskine er for hurtig til at vise forskelle, så øg antallet af records med en faktor 10.

¹ Jeg har støttet mig til [https://technet.microsoft.com/en-us/library/ms188609\(v=sql.110\).aspx](https://technet.microsoft.com/en-us/library/ms188609(v=sql.110).aspx)

² <https://docs.microsoft.com/en-us/sql/integration-services/import-export-data/import-and-export-data-with-the-sql-server-import-and-export-wizard>

³ <https://docs.microsoft.com/en-us/sql/relational-databases/indexes/clustered-and-nonclustered-indexes-described>