

Korszerű adatbázisok előadás 03



Miről lesz szó?

Lekérdezések optimalizálása

- Indexek
- Join-ok
- Eszközök
 - Lekérdezés végrehajtási terv
 - SQL Server Profiler
 - Tuning Advisor



Hogyan lehet gyorsítani a lekérdezések sebességét?

Sok esetben az a probléma, hogy keresésnél akár az összes rekordot végig kell nézni (TABLE SCAN)

```
■SELECT
   FROM noindex rendeles tetel
   WHERE EGYSEGAR > 500
100 % ▼ ⊲
Query 1: Query cost (relative to the batch): 100%
       FROM [noindex rendeles tetel] WHERE [EGYSEGAR]>@1
                Table Scan
           [noindex rendeles tetel]
 SELECT
               Cost: 100 %
Cost: 0
                 0.116s
                38332 (99%)
```

ÖTLET1: Rendezzük fizikailag sorrendbe az adatokat!



ÓBUDA UNIVERSITY Mi a baj a fizikai rendezéssel?

Sok adatot kell mozgatni

Vagy növekvő, vagy csökkenő

Csak egyféle rendezési szempont

DML utasításoknál újra kell rendezni

DML = Data Manipulation Language (Adatok beszúrása, törlése, módosítása)



ÓBUDA UNIVERSITY Mi a baj a fizikai rendezéssel?

RAKTAR _KOD	RAKTAR_NEV	RAKTAR_CIM
7	Báthory	1152 Budapest Báthory u. 2.
8	Havanna	1181 Budapest Havanna u. 3
4	KIS	1039 Budapest Madzsar J. u. 23.
1	Kisfaludy	2135 Csörög Kisfaludy u. 55.
5	NAGY	2089 Telki Nefelejcs u. 27
2	Nagyszeben	1182 Budapest Nagyszeben u. 12.
3	Toldy	1015 Budapest Toldy F. u. 58.
6	Veresegyházi	2112 Veresegyház Új iskola u. 10.

INSERT INTO Raktar VALUES (9, 'Lapos', '1016 Budapest Krisztina krt')

ÖTLET2: Rendezzük logikailag sorrendbe az adatokat!



Mit jelent a logikai rendezés?

Csak a rendezés alapjául szolgáló mezőt (vagy kifejezést) és a rekordok azonosítóját (mutató, memóriacím) tároljuk

- Az adatok eredeti tárolási sorrendje nem változik
- Egy táblához több logikai rendezést is létrehozhatunk
- A logikai rendezést indexelésnek is nevezik

Indexek

Az index a táblához vagy nézethez rendelt olyan speciális adatstruktúra, amely felgyorsítja a lekérdezések sebességét.

INDEX (Név szerint)					
Név	ID				
Bódi István	D02	\			
Fehér Katalin	D04	_			
Kiss Béla	D01	/			
Nagy Ilona	D03				

DOLGOZÓ					
ID	Név	Életkor			
D01	Kiss Béla	22			
D02	Bódi István	18			
D03	Nagy Ilona	32			
D04	Fehér Katalin	18			



Index - példa

```
Columns
       SORSZAM (PK, FK, int, not null)
       TERMEKKOD (PK, FK, nvarchar(255), nc
       ■ EGYSEGAR (float, null)

■ MENNYISEG (float, not null)

      Keys
       PK Rendeles tetel
       FK_Rendeles_tetel_Rendeles
       FK_Rendeles_tetel_Termek
       Constraints
       Triggers
       Indexes
       PK_Rendeles_tetel (Clustered)
```

```
■SELECT
   FROM rendeles tetel
   WHERE EGYSEGAR > 500
100 % ▼ <
Ouery 1: Ouery cost (relative to the batch): 100%
SELECT * FROM [rendeles tetel] WHERE [EGYSEGAR]>@1
            Clustered Index Scan (Cluste...
            [Rendeles tetel].[PK Rendele ...
 SELECT
                   Cost: 100 %
Cost: 0 '
                    0.057s
                    38280 of
                  38149 (100%)
```

A Rendeles_tetel táblához indexet készítve a végrehajtási idő töredékére csökken

Indexek csoportosítása

Egyedi 🗲 ⋺ Duplikált

• Egy index érték csak egyszer fordulhat-e elő?

Sűrű ← → Ritka

Minden adatrekordhoz készül index bejegyzés?

Egyszerű 🗲 ⋺ Összetett

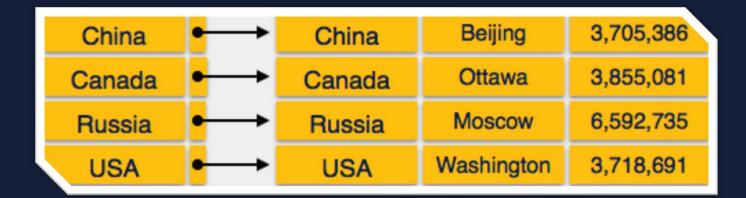
• Egy vagy több mezőre épül-e?

Növekvő ← → Csökkenő

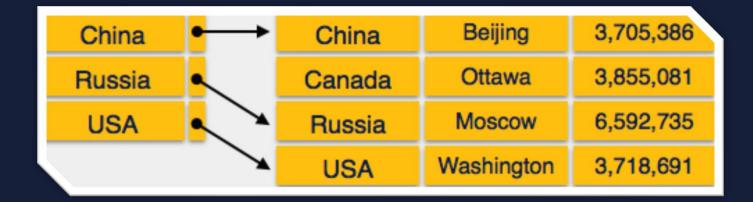
Milyen irányú a rendezés?



Sűrű vs. Ritka index



Az indexmutató egy rekordra mutat (sűrű index).



Az indexmutató egy blokkra mutat (ritka index). A blokkon belül a keresés szekvenciális

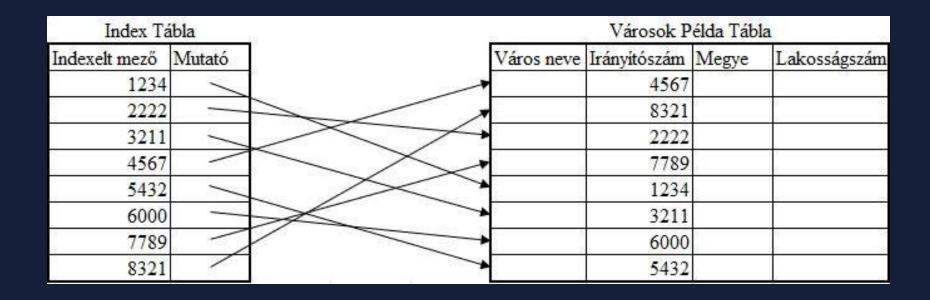


Index adatstruktúrák

- Egyszintes indexek
- Többszintes indexek
- B-fák
- Hash-alapú indexek
- Bitmap indexek

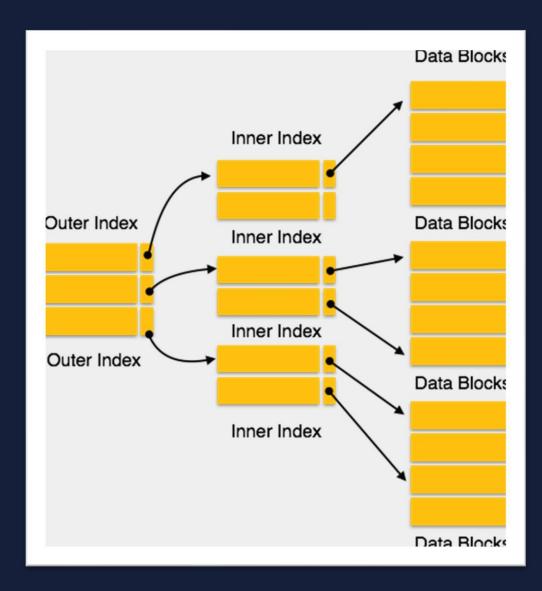
Egyszintes indexek

Két mezőből álló indextábla, amely az indexelt mező alapján sorba van rendezve. A mutató a rekord fizikai helyére mutat.





Többszintes indexek

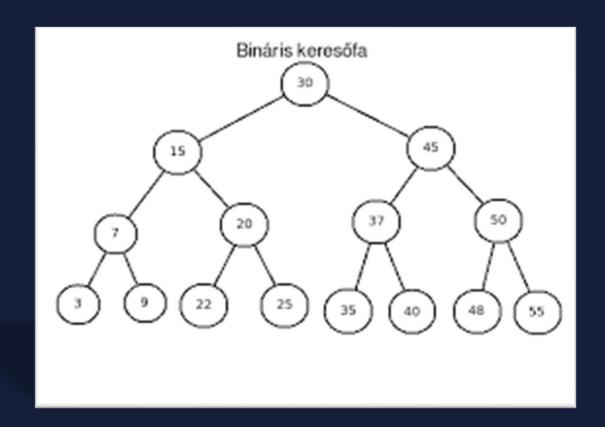


Az indexekhez is indexet készítünk

- Az index így kisebb részekből áll
- Hasznos, ha az egyszintes index nem fér el a memóriában
- Kevesebb blokk olvasás szükséges az adat megtalálásához



Keresőfák



Bármely csomópontból kiindulva a csomópont bal részfájában csak a csomópontban elhelyezettnél kisebb, a jobb részfájában pedig csak nagyobb értékek szerepelnek

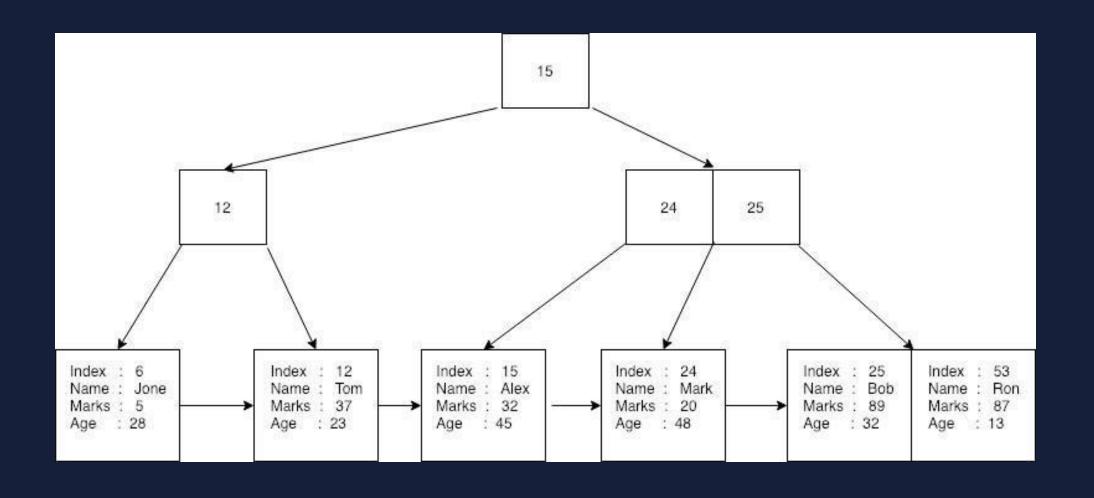


B-fák tulajdonságai

- A gyökértől a levelekig vezető utak hossza egyforma
- Az indexek a B-fa csomópontjaiban helyezkednek el
- Az adatok helyét jelző mutató csak a levelekben található
- A struktúra lehetővé teszi a soros és a random elérést is

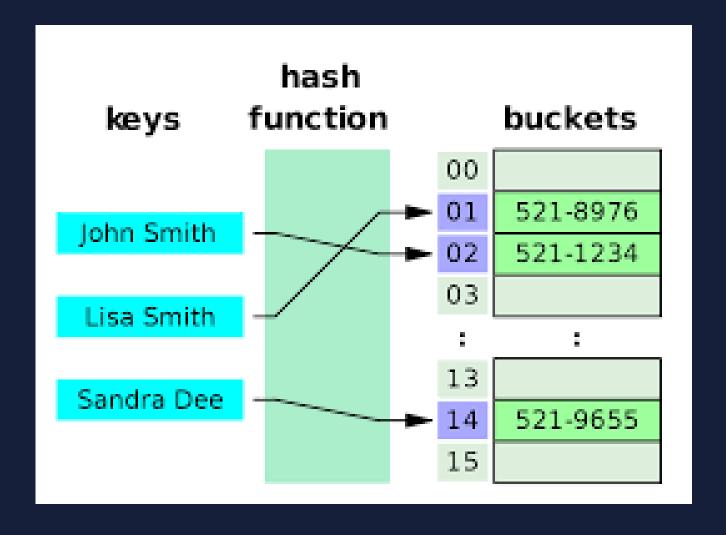


ÓBUDA UNIVERSITY B-fák (kiegyensúlyozott keresőfák)





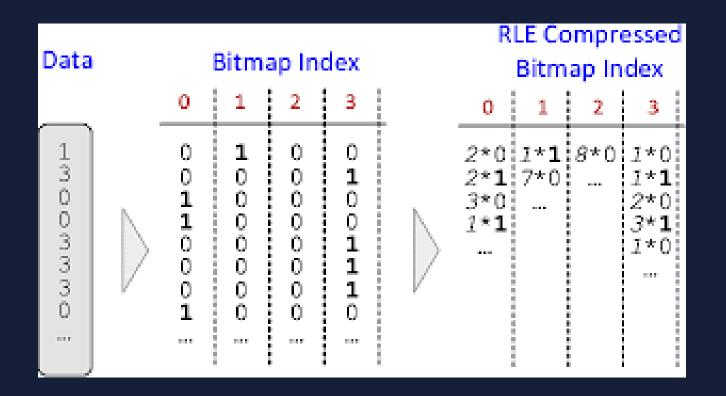
Hash-alapú indexek



- Az adatok csoportokba vannak rendezve
- A hash függvény adja meg, hogy melyik csoportban van az adat



Bitmap indexek



- Olyan oszlopokra alkalmazzuk, ahol kevés az egyedi érték
- Az index tömöríthető is



ÓBUDA UNIVERSITY Fontosabb T-SQL Index típusok

- Clustered
- Non-clustered
- Columnstore
- XML
- Spatial

Az indexek létrejöhetnek automatikusan vagy manuálisan (CREATE INDEX)



OE OBUDA UNIVERSITY Clustered index

- Az adatokat az index kulcsnak megfelelő sorrendbe rendezi és tárolja.
- A clustered index B-fa struktúrát használ
- Egy tábla esetén csak egy clustered index hozható létre
- Alapértelmezés szerint az elsődleges kulcs definiálásakor automatikusan létrejön

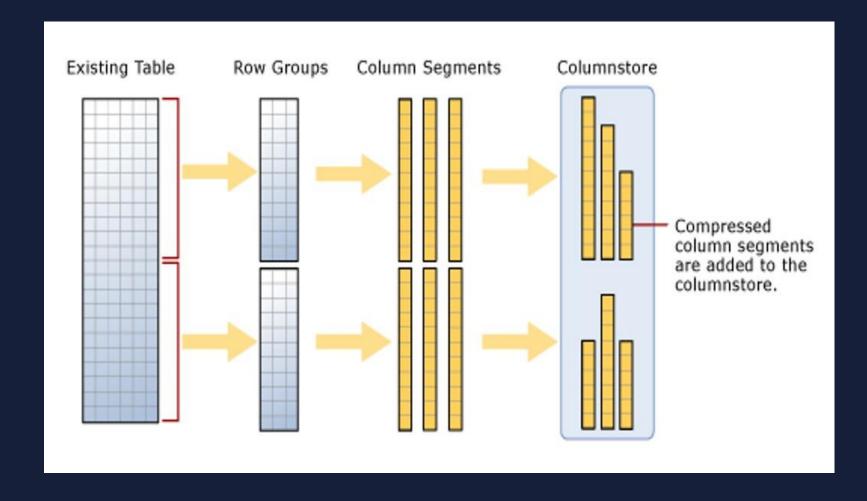


ÓBUDA UNIVERSITY Non-clustered index

- Kulcs-mutató érték párokat tárol.
- Az index sorai a kulcs értékeknek megfelelő sorrendben vannak tárolva
- Az adatok tárolási sorrendje ettől eltérő
- Egy tábla esetén több non-clustered index is létrehozható
- Speciális lehetőségek: Unique, Filtered, Included columns



Columnstore index



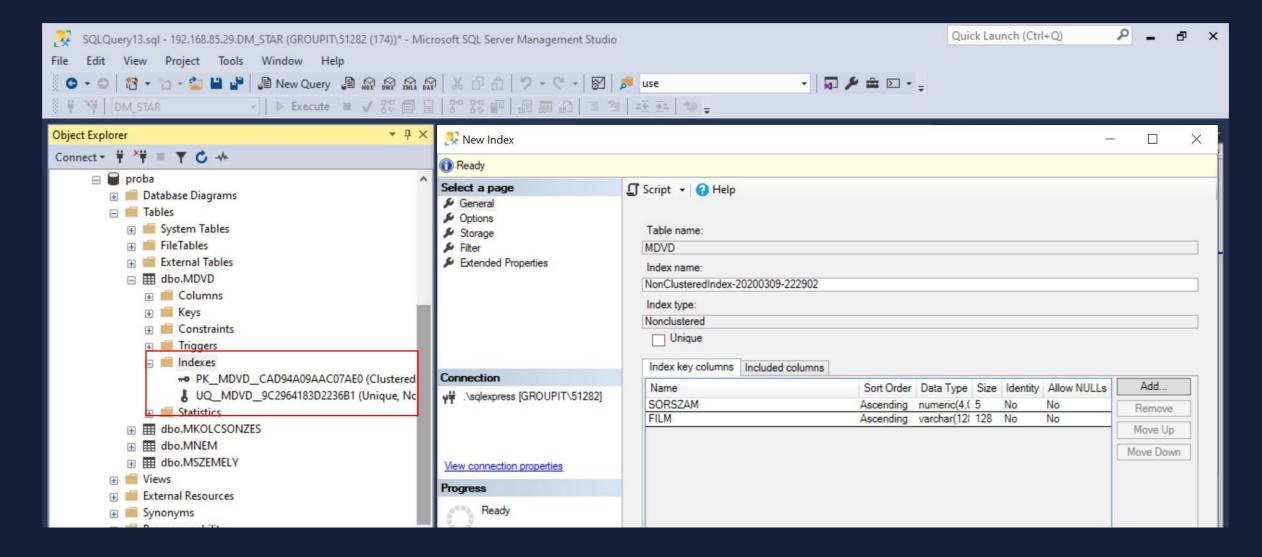
- A columnstore index lehet
 - Clustered
 - Non-clustered
- Egy táblához csak egy columnstore index készíthető

Adattárházakból való lekérdezéseknél kiemelten fontos!

Oszlop-alapú tárolás és lekérdezés végrehajtás



OBUDA UNIVERSITY INCEXEK - SSMS





OBUDA UNIVERSITY Index létrehozása paranccsal – példa1

CREATE [[UNIQUE][CLUSTERED][COLUMNSTORE]] INDEX indexnév ON táblanév (oszlop1 [ASC] [DESC], oszlop2 [ASC] [DESC] ...) [INCLUDE (oszlopok)]

```
SQLQuery8.sql - hal...hop (hallgato (87))* 

□ × SQLQuery5.sql - hal...hop (hallgato (93))*
                                                     SQLQuery3.sql - hal...hop (hallgato (70
  --- Nonclustered Unique index a Raktár táblára
       raktárkód szerint növekvő
  PCREATE UNIQUE INDEX I Raktar ON Raktar
         RAKTAR KOD ASC
```

UNIQUE kényszer létrehozásakor a megfelelő UNIQUE index automatikusan létrejön!



ÓBUDA UNIVERSITY Index létrehozása paranccsal – példa2

CREATE [[UNIQUE][**CLUSTERED**][COLUMNSTORE]] **INDEX** indexnév **ON** táblanév (oszlop1 [ASC]|[DESC], oszlop2 [ASC]|[DESC] ...)) [INCLUDE (oszlopok)]

```
SQLQuery11.sql - h...bshop (ujaenb (89))* 

SQLQuery8.sql - hal...hop (hallgato (87))*

SQLQuery5.sql - hal...hop (hallgato (93))*
                                                                             SQLQuery
  --- Clustered index a noindex rendeles táblára
   -- sorszam szerint növekvő
  PCREATE CLUSTERED INDEX I rendeles ON noindex rendeles
         SORSZAM ASC
```

Elsődleges kulcs létrehozásakor a Clustered Index automatikusan létrejön!



oe obuda university Index létrehozása paranccsal – példa3

CREATE [[UNIQUE][CLUSTERED][**COLUMNSTORE**]] **INDEX** indexnév **ON** táblanév (oszlop1 [ASC]|[DESC], oszlop2 [ASC]|[DESC] ...)) [INCLUDE (oszlopok)]

```
SQLQuery13.sql -...BKAE\gmolnar (61))* → × SQLQuery12.sql - h...bshop (ujaenb (89))* SQLQuery2.sql - hal...hop (hallgato (58))*
  ---COLUMNSTORE Index létrehozása az Ugyfel
   --tábla LOGIN és EMAIL oszlopaival
  PCREATE COLUMNSTORE INDEX i login email_ugyfel
   ON Ugyfel (LOGIN, EMAIL)
```

A Clustered index létrehozását az Azure SQL nem minden szolgáltatásrétege támogatja



Melyik index legyen használva?

HINT- Lekérdezési tipp

Használata:

SELECT

FROM...

•••

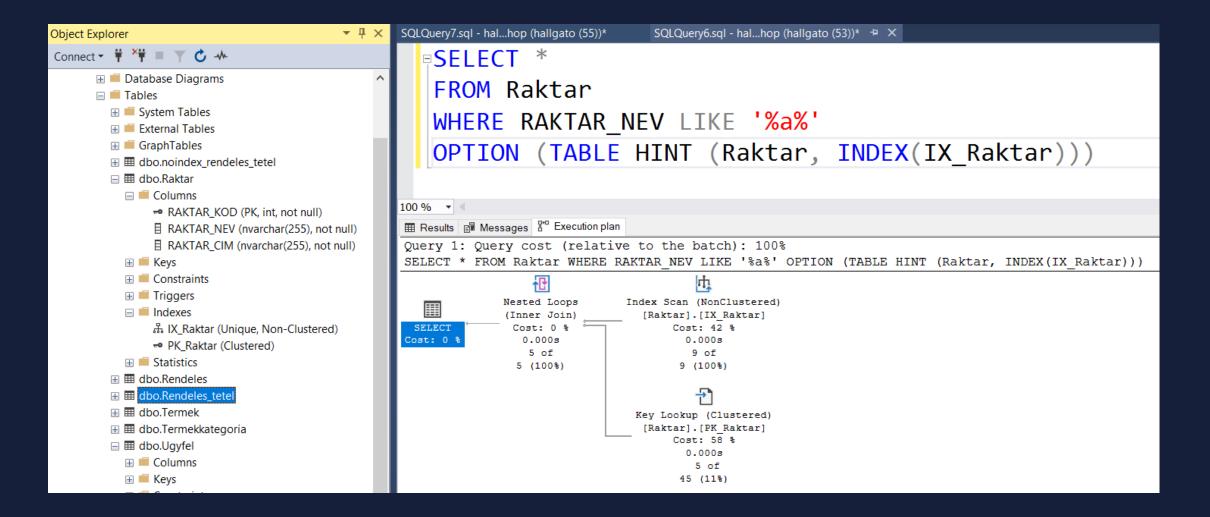
ORDER BY ...

OPTION (TABLE HINT (táblanév, INDEX(indexnév)))

- Többféle HINT létezik:
 - JOIN HINT
 - TABLE HINT
 - QUERY HINT
- Céljuk:
 - Preferált vagy tiltott index
 - JOIN típus, sorrend
 - Táblaelérési mód
 - Párhuzamos végrehajtás

Az SQL Server Query Optimizer által alkalmazott beállításokat csak indokolt esetben bíráljuk felül a HINT-ekkel

ÓBUDA UNIVERSITY HINT példa





Indexek hátrányai

- Tárhelyet foglalnak
- Folyamatos karbantartást igényelnek
 - Minden DML-művelet esetén (automatikus)
 - Újraépítés (Rebuild) törli és újból létrehozza az indexet
 - Újraszervezés (Reorganize) töredezettségmentesítést hajt végre
 - Tömörítés
 - Index statisztikák a lekérdezések optimalizálását segítik
- A DML-műveleteket lelassítják
- Nem ajánlottak
 - Kis táblák esetén
 - Sok NULL értéket tartalmazó oszlopra
 - Olyan oszlopokra, amelynek értékei gyakran változnak



ÓBUDAI EGYETEM Táblák ÖSSZEKapcsolása

- Akkor lehet rá szükség, ha a keresett információ egynél több táblában található meg
- A kapcsolat alapja legtöbbször az egyik tábla idegen kulcsának és a másik tábla kulcsának azonos értéke, ritkább esetben a két tábla egyegy oszlopára vonatkozó összehasonlító feltétel
- Az összekapcsolás megvalósításának tipikus módja:

SELECT oszlopnév lista

FROM táblanév1 aliasnév1 INNER JOIN táblanév2 aliasnév2 **ON** aliasnév1.idegenkulcsérték = aliasnév2.kulcsérték



Táblák összekapcsolása - Megjegyzések

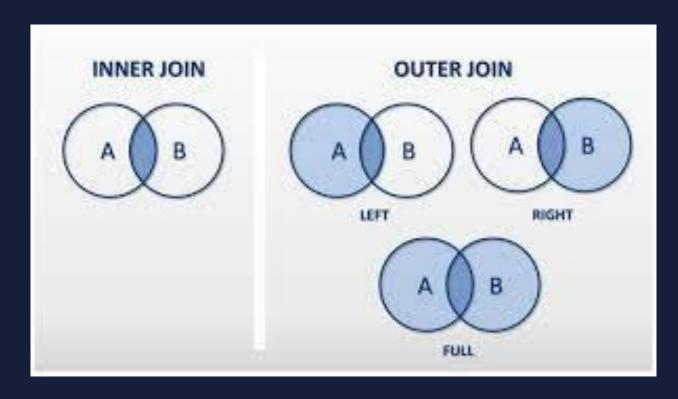
- Az aliasnevek (másodlagos táblanevek, rövid nevek) használata javítja a lekérdezés olvashatóságát*
- Egy tábla akár saját magával is összekapcsolható, ilyenkor az aliasnevek használata kötelező
- Mivel a két táblának lehetnek azonos nevű mezői is, ezért a SELECT utáni oszlopnév listában használjunk minősített oszlopneveket (tábla aliasnév.oszlopnév) vagy aliasnevek hiányában táblanév.oszlopnév formában
- Az INNER kulcsszó használata nem kötelező
- Az ON kulcsszó után összetett feltétel is szerepelhet
- Az eredménylistából INNER JOIN esetén kimaradnak azok a sorok, ahol az idegenkulcs
 kulcs feltétel valamelyik oldalán NULL érték található, pl: a diák azonosítója nincs megadva a munka táblában

^{*}Egy tábla neve a legáltalánosabb esetben [Szervernév].[Adatbázisnév].[Sémanév].[Táblanév] formátumú, nagyon hosszú is lehet



Fontosabb JOIN típusok

- (INNER) JOIN:
 Az A tábla idegen kulcsa megegyezik a B tábla kulcsával
- LEFT (OUTER) JOIN:
 Az INNER JOIN eredményéhez hozzá veszi az A tábla minden további sorát is
- RIGHT (OUTER) JOIN:
 Az INNER JOIN eredményéhez hozzá veszi az B tábla minden további sorát is
- FULL (OUTER) JOINT:
 Az INNER JOIN eredményéhez hozzá veszi az A és B tábla minden további sorait is



Az OUTER szó használata nem kötelező

Az első három esetben különbség akkor van, ha az idegenkulcs = kulcs feltétel egyik oldalán NULL érték áll

CROSS JOIN

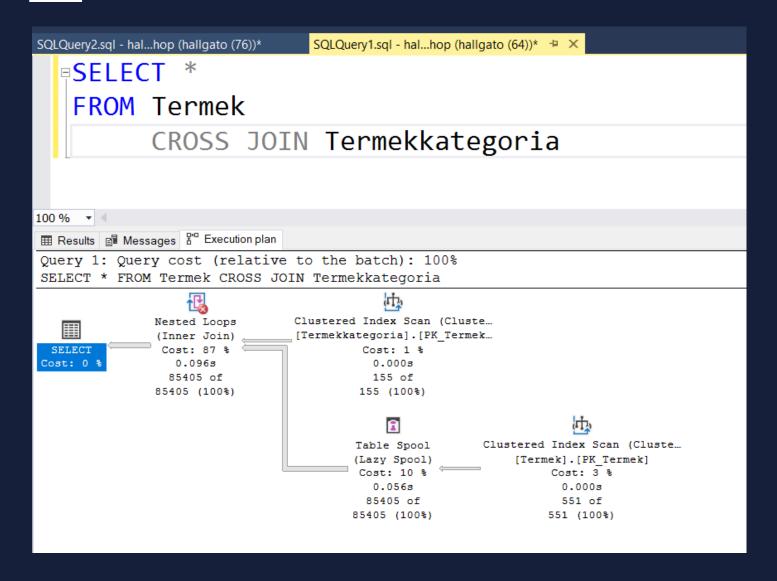
Az első tábla minden egyes sorát összepárosítja a második tábla minden egyes sorával (Descartes-szorzat)

- Kétféle szintaktika
 - SELECT ...
 FROM Tábla1 CROSS JOIN Tábla2
 - SELECT...
 FROM Tábla1, Tábla2

A CROSS JOIN-os lekérdezések nagyon költségesek lehetnek!



CROSS JOIN - Példa



Több, mint 85 ezer sor lesz az eredmény!

(INNER JOIN esetén csak 551 keletkezik)

JOIN HINT-ek

A táblák összekapcsolását finomhangolják

- Fontosabb HINT-ek
 - MERGE
 - HASH
 - LOOP*
 - REMOTE**

Formája

SELECT *

FROM Tábla1 t1

[INNER|OUTER|FULL] [HINT] JOIN Tábla2 t2

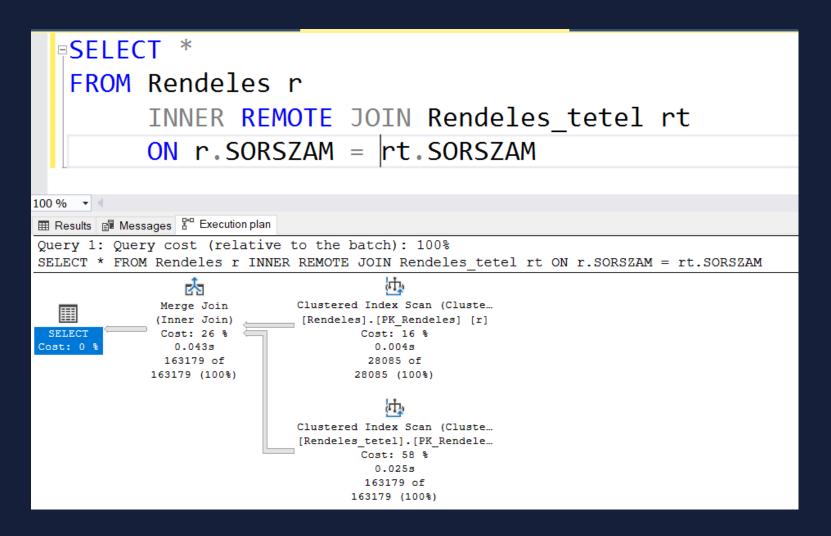
ON ...

^{*} A LOOP HINT nem használható RIGHT és FULL JOIN esetén

^{**}A REMOTE HINT csak INNER JOIN esetén használható, nem minden verzió támogatja



JOIN HINT - példa



A REMOTE HINT-et akkor érdemes használni, ha

 A jobboldali tábla egy távoli adatforrásban van

ÉS

 Nagyobb méretű, mint a baloldali (helyi) tábla

APPLY operátor

A JOIN alternatívája, táblaértékű kifejezéseket kapcsol össze

- Két fajtája:
 - CROSS APPLY: a baloldali tábla kifejezésből azokat a sorokat adja vissza, amelyekkel van egyező a jobboldali tábla kifejezésben (~ INNER JOIN)
 - OUTER APPLY: a baloldali tábla kifejezés minden sorát visszaadja, azoknál a soroknál, ahol nincs megfelelő a jobboldali tábla kifejezésben, ott a jobboldali értékeknél NULL értékeket ad vissza (~ LEFT JOIN)

APPLY formája

```
SELECT ...
  FROM tábla1 t1
  CROSS | OUTER APPLY
   SELECT ...
   FROM tábla2 t2
   WHERE t1.oszlop1 = t2.oszlop2
  ) AS t
```

Az APPLY operátorok használata esetén a lekérdezés végrehajtási terve más lesz, mint JOIN-ok esetén.



APPLY vs. JOIN

- Az APPLY operátor esetén a kapcsolat jobb oldala egy tábla értékű függvény is lehet
- Bizonyos esetekben az APPLY operátor használata performanciálisan kedvezőbb lehet a JOIN-nál
- Az APPLY T-SQL specifikus operátor, a JOIN viszont szabványos



APPLY példák

```
SQLQuery6.sql - hal...hop (hallgato (54))* + X SQLQuery5.sql - not connected*

        □SELECT r.rend_datum,

               u2.nev.
               r.fiz_mod
   FROM Rendeles r
   CROSS APPLY
     SELECT LOGIN, nev
     FROM Ugyfel u
     WHERE r.LOGIN = u.LOGIN
      AS u2
100 % ▼
Query 1: Query cost (relative to the batch): 100%
SELECT r.rend datum, u2.nev, r.fiz mod FROM Rendeles r CROSS APPLY
Missing Index (Impact 97.5938): CREATE NONCLUSTERED INDEX [<Name of
              Hash Match
                            Clustered Index Scan (Cluste ...
             (Inner Join)
                              [Ugyfel].[PK Ugyfel] [u]
             Cost: 49 % ==
Cost: 0
               0.023s
                                     0.000s
              28085 of
                                     200 of
             28085 (100%)
                                    200 (100%)
                            Clustered Index Scan (Cluste...
                             [Rendeles].[PK Rendeles] [r]
                                    Cost: 50 %
                                     0.013s
```

```
SELECT r.rend datum,

               u2.nev,
               r.fiz mod
   FROM Rendeles
   OUTER APPLY
       dbo.GetAllClient() AS u2
   WHERE YEAR(r.rend_datum)>2016
100 % ▼ 4
Query 1: Query cost (relative to the batch): 100%
SELECT r.rend datum, u2.nev, r.fiz mod FROM Rendeles r OUTER APPLY dbo.GetAllClien
               Nested Loops
                                Clustered Index Scan (Cluste ...
             (Left Outer Join)
                                 [Rendeles].[PK Rendeles] [r]
               Cost: 79 %
                                        Cost: 8 %
Cost: 0
                 0.903s
                                        0.019s
                713000 of
                                        3565 of
               717800 (99%)
                                       3589 (99%)
                                       Table Spool
                                                     Clustered Index Scan (Cluste...
                                       (Lazy Spool)
                                                        [Ugyfel].[PK Ugyfel]
                                       Cost: 13 %
                                                            Cost: 0 %
                                         0.471s
                                                             0.000s
                                        713000 of
                                                             200 of
                                      717800 (99%)
                                                            200 (100%)
```

Lekérdezés optimalizáló - automatikus

Cél a lekérdezés által használt hardver erőforrásigény (CPU, Memória, I/O) csökkentése

Feladatai:

- Fizikai terv elkészítése
- Táblák bejárása
- Táblák összekapcsolása
- Logikai terv
- Optimalizálás (költségbecslés alapján)



Lekérdezés optimalizáló - optimalizálás

Gyakran alkalmazott technikák

- Statisztikák
- Táblaelérési módok
- Indexek
- Hint-ek
- Plan cache itt tárolja el azokat a végrehajtási terveket, amelyek számításba jöhetnek



ÓBUDAI EGYETEM Lekérdezés optimalizálás manuálisan

- Egy lekérdezés általában sokféle módon végrehajtható
- A beépített lekérdezés optimalizáló (SQL Query Optimizer) nem mindig találja meg az optimális módot
 - Ez főleg komplex lekérdezéseknél fordulhat elő
 - Ilyenkor nem vizsgál meg minden lehetséges esetet, mert akkor túl sokáig tartana az optimalizálási folyamat
 - Elsősorban azt próbálja megbecsülni, hogy az adott terv végrehajtása hány sort érinthet
 - Gyakori futtatás esetén sokszor a már meglévő tervet használja

Nem megfelelő performancia esetén érdemes felülvizsgálni a végrehajtási tervet



Lekérdezések optimalizálása – néhány javaslat

- SELECT * helyett jobb a SELECT oszloplista
- Lehetőleg kerüljük a DISTINCT használatát
- Helyettesítő karaktereket lehetőleg csak a keresett kifejezés végén használjunk
- A HAVING részben csak aggregált adatot szűrjünk
- Beágyazott SELECT-nél az IN helyett inkább EXISTS operátort használjunk
- UNION helyett nagy tábláknál inkább UNION ALL-t használjunk
- Ha a WHERE feltételben függvényt használunk, akkor az lehetőleg az operátor jobb oldalán legyen



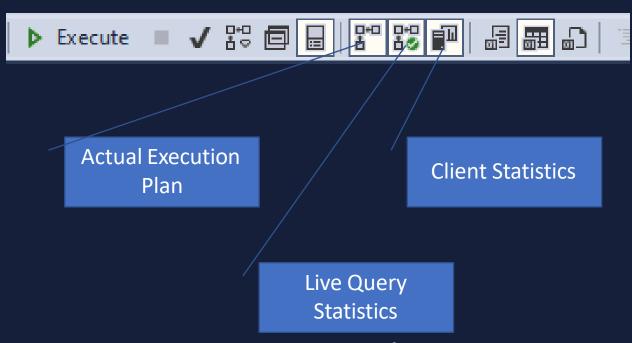
Lekérdezések optimalizálása – néhány javaslat

- A JOIN és a WHERE részekben kerüljük a számított mezők használatát
- Részesítsük előnyben az analitikus függvények használatát
- A GROUP BY oszloplista sorrendjét ha lehet úgy válasszuk meg, hogy az elején egyedi értékeket tartalmazó oszlopok legyenek
- A hosszú táblanevek helyett használjunk rövid alias neveket
- A WHERE részben a > jel helyett általában előnyösebb a >=
- Ha a WHERE részben több feltétel van, akkor azt vegyül előre, amely ahhoz a táblához tartozik, amelytől a legtöbb más tábla függ
- A FROM részben pont fordítva, itt a legtöbb függéssel rendelkező tábla legyen az utolsó



Lekérdezések optimalizálása – eszközök

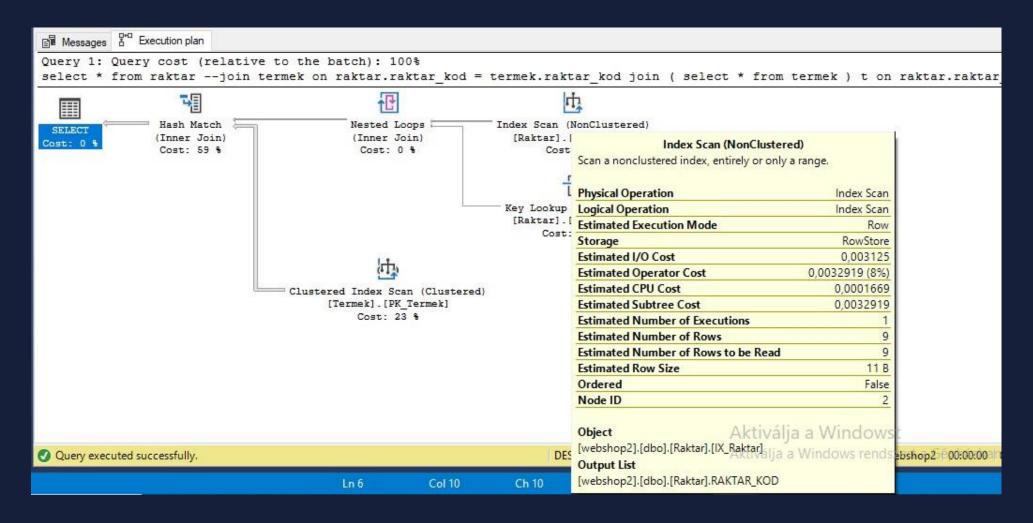
- Lekérdezés végrehajtási terv
- SQL Tuning Advisor
- SQL Server Exended Events
- SQL Profiler, XEvent Profiler
- Egyéb



- SET STATISTICS IO ON/OFF, SET STATISTICS TIME ON/OFF
- SET SHOWPLAN_TEXT ON/OF, SET SHOWPLAN_ALL ON/OFF
- Dynamics Management Objects



Lekérdezés végrehajtási terv



SSMS-ben: Query\Display Estimated Execution Plan

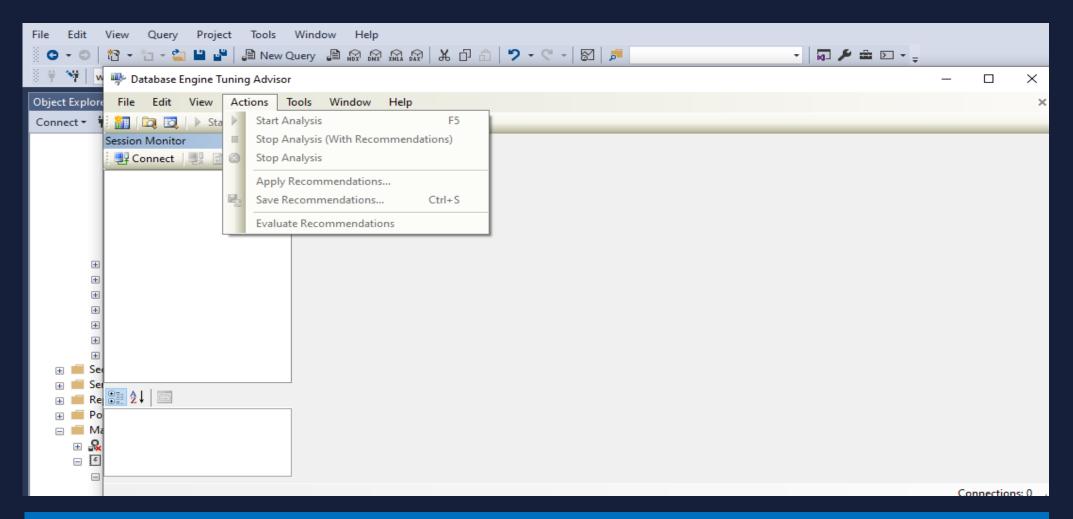


Lekérdezés végrehajtási terv – gyakori operátorok

1	Table Scan	Scan of a whole table stored as a heap. A table can be organized as a heap or as a clustered index.
<u> </u>	Clustered Index Scan	Scan of a whole table stored as a clustered index. Indexes are stored as balanced trees.
	Clustered Index Seek	SQL Server seeks for the first value in the seek argument (for example, a column value in the WHERE clause) in a clustered index and then performs a partial scan.
蟲	Index Scan	Scan of a whole nonclustered index.
क्षे	Index Seek	SQL Server seeks for the first value in the seek argument (for example, a column value in the WHERE clause) in a nonclustered index and then performs a partial scan.
Sept.	RID Lookup	Lookup for a single row in a table stored as a heap by using its row identifier (RID).
1010 1010 10101 01010	Key Lookup	Lookup for a single row in a table stored as a clustered index by using the key of the index.
-	Hash Match Join	Joins that use the Hash algorithm.
	Merge Join	Joins that use the Merge algorithm.
10	Nested Loops	Joins that use the Nested Loops algorithm.
Σ Σ Σ Σ	Stream Aggregate	Aggregation of ordered rows.
	Hash Match Aggregate	Hash algorithm used for aggregating. Note that the icon is the same as the icon for the Hash Match Join; however, in an execution plan, text below the icons gives you information about whether the operator performed a join or an aggregate.
7	Filter	Filters rows based on a predicate (for example, a predicate of the WHERE clause).
21	Sort	Sort of incoming rows.



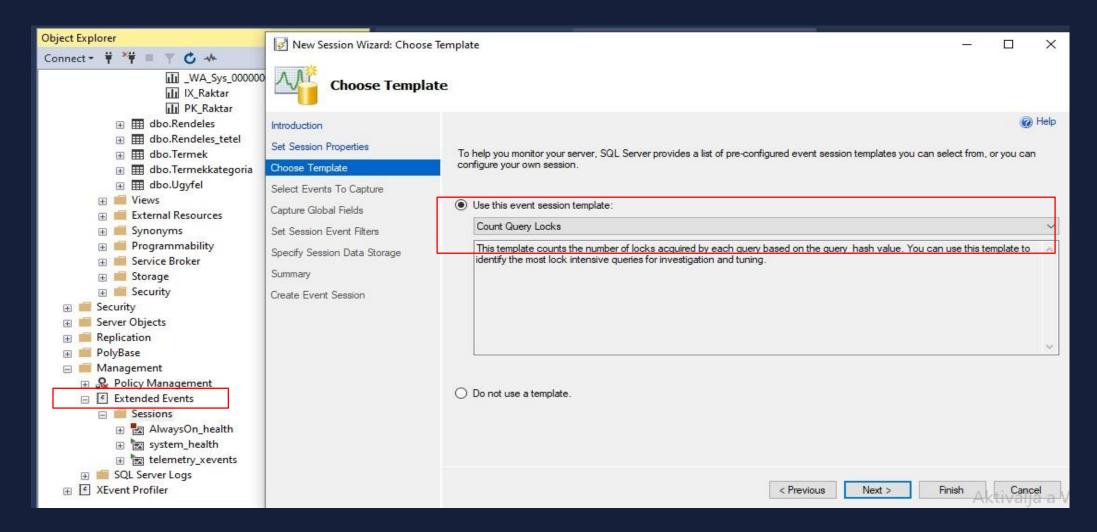
Tuning Advisor



Az SQL Server Express és az Azure SQL ezt a funkciót nem támogatja. Opcionálisan más verzió vagy az Azure Portál teljesítménykezelő funkciói használhatók.



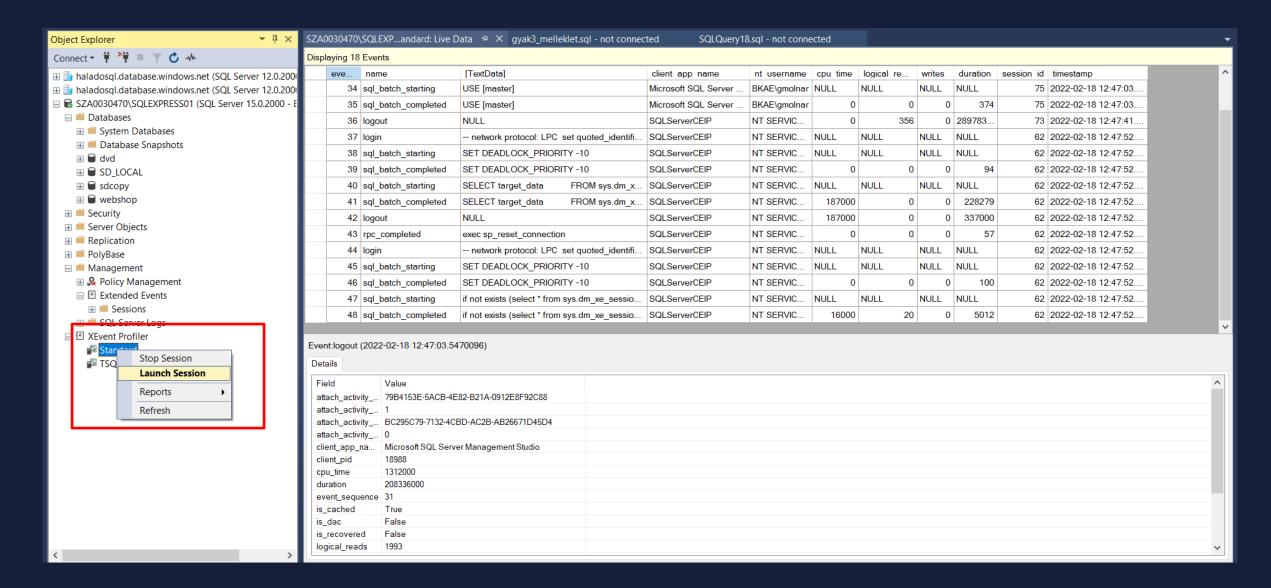
SQL Server Extended Events



Az Extended Events futtatásához ALTER ANY EVENT SESSION jogosultság szükséges.

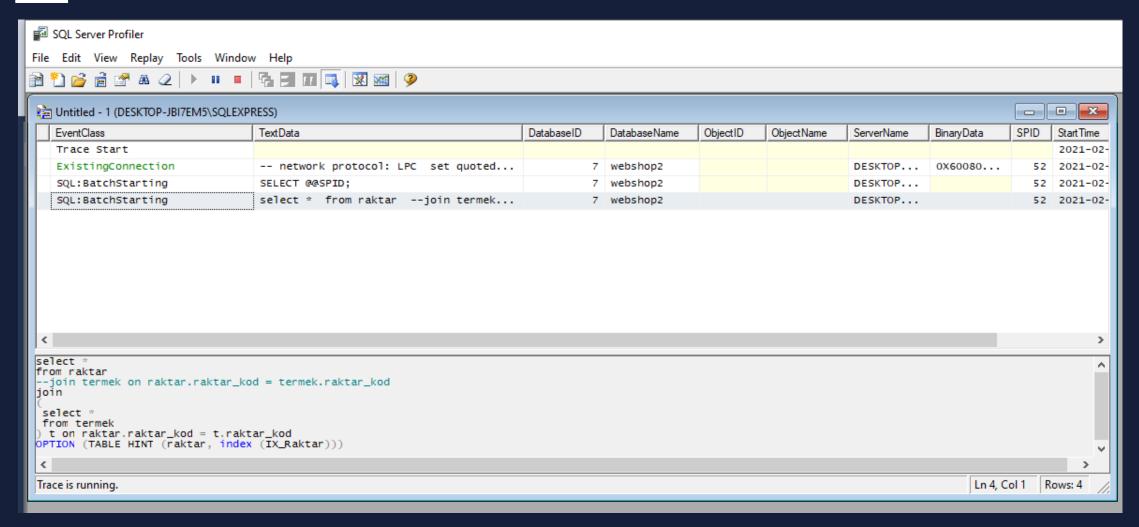


SQL Server XEvent Profiler





SQL Server Profiler



Az Azure SQL ezt a szolgáltatást nem támogatja. Opcionálisan az Azure Data Studio Profiler Extension használható helyette