Microsoft SQL (MS BI) (Training 360)
1. forduló

Ismertető a feladathoz

Az UnCovid Zrt-nél dolgozol junior MS SQL Server BI fejlesztőként. A cég nemrég bővítette a profilját, és most már a koronavírussal kapcsolatos egészségügyi termékeket is forgalmazza. A cég kereskedelmi rendszere MS SQL Server 2019 Enterprise Edition alapú, a webáruház PHP/MySQL alapú, a könyvelési rendszer pedig Oracle adatbázist használ. Mindemellett kiépítésre kerül egy adatgyűjtő rendszer, amelyik a regionális, országos és globális, COVID-19-kel kapcsolatos esetek elérhető adatait gyűjti össze egy SQL Server 2019 adattbázisba, és gondoskodik a tisztításukról is. A heterogén forrásból származó és nagy adatmennyiségű (>10TB) adatok miatt adattárházat kell építeni, és naponta frissíteni a különféle rendszerekből érkező adatokkal. Az adattárház adatai alapján multidimenzionális adatmodellt kell kialakítani MS SQL Server OLAP adatbázis formájában, melyet az üzleti elemzők fognak használni Power BI-ban.

Junior fejlesztőként mások által készített Integration Services (SSIS), Analysis Services (SSAS) és Reporting Services (SSRS) projektekben kell apró módosításokat végezned. Ha a BI fejlesztési csapat vezetője elégedett lesz a munkáddal (azaz az első két szintet sikeresen teljesíted), akkor önálló csomagok, kockák és riportok írását is Rád bízzák. Ha ezekben is jeleskedsz (azaz a következő három szintet is teljesíted), akkor önálló projektek megtervezésében is részt tudsz venni. A legutolsó két szinten azt kell bizonyítanod, hogy képes vagy komplett, adattárház-alapú rendszerek megtervezésére és elkészítésére is.

Tekintettel arra, hogy egy választ sem rögzítettél az alábbi feladatlapon, ebben a fordulóban a kitöltésére rendelkezésre álló idő teljes egésze, azaz 10 perc került rögzítésre mint megoldáshoz felhasznált idő.

1. feladat 0 / 1 pont

Egy szenior BI fejlesztő asszisztenseként azt a feladatot kaptad, hogy tanulmányozd át az eddig elkészített SQL Server Integration Services csomagokat. Éppen akkor szakadozott a Teams-ben a hang, amikor a telepítési modellekről beszélt a szenior BI fejlesztő. Milyen telepítési modellek léteznek az SSIS-ben?

- [ ] Task
- [x] Project
- [ ] Flow
- [ ] Program
- [ ] Process
- [x] Package
- [ ] Component

Magyarázat a megoldáshoz

Az SQL Server Integration Services kétféle telepítési platformot támogat: Package és Project. A 2012-es verzió előtt csak Package telepítés létezett. A 2012-es verzió ezt egészítette ki egy komplett SSIS projekt telepítésének lehetőségével.

https://docs.microsoft.com/en-us/sql/integration-services/packages/deploy-integration-services-ssis-projects-and-packages?view=sql-server-ver15

2. feladat 0 / 8 pont

Sikeresen megtaláltad a csomagokat a gépeden, és az első csomag legelső Control Flow doboza egy For Each Loop konténer. Jelöld meg az igaz kijelentéseket a For Each Loop konténerre vonatkozóan!

- [x] A konténeren belüli taszkok közös változókat használhatnak
- [ ] A konténer szerepe az, hogy a benne szereplő taszkok csak egymás után legyenek végrehajthatók.
- [ ] A konténerben szereplő taszkok egy változóban megadott szám alapján ismétlődnek.
- [x] A konténerben szereplő taszkokhoz közös eseménykezelő eljárást készíthetünk.
- [ ] A konténer leginkább arra való, hogy egy kurzorral végig olvassuk egy tábla sorait.

Magyarázat a megoldáshoz

Ez a kérdés a *For Each Loop* konténert célozta be. Nem igaz az, hogy a konténerben lévő taszkok csak egymás után hajthatók végre – kivéve, ha előfeltétel megszorítást alkalmazunk közöttük. A *For Each Loop* konténerre nem igaz (a *For Loop*-ra igaz lenne), hogy a taszkok egy előre definiált változó szerinti darabszámban hajtódnak végre. A közös változókezelésre és a közös eseménykezelésre vonatkozó kitételek viszont igazak.

https://docs.microsoft.com/en-us/sql/integration-services/control-flow/foreach-loop-container?view=sql-server-2017

3. feladat 0 / 7 pont

A BI csapat és az üzleti felhasználók közötti egyeztető megbeszélésen azt a feladatot kapod, hogy a beszélgetés során feltárt üzleti objektumokat két külön oszlopba rendezd. Az első oszlopba a dimenziók, a másodikba a tényobjektumok kerüljenek. Az alábbi felsorolásból mely objektumok kerülnek az első oszlopba?

- [ ] Napi új fertőzöttek adatai egy lokációban
- [ ] Termék beérkezések egy raktárba
- [ ] Visszaigazolt vevői árurendelések
- [x] Földrajzi egységek
- [ ] Termékek aktuális készletadatai
- [ ] Aktuális devizaárfolyamok
- [x] Raktárak

Magyarázat a megoldáshoz

A ténytáblák a megtörtént tranzakciók adatait tartalmazzák, amelyeket a dimenziótáblák adatai alapján tudunk elemezni. Mindezek alapján egyedül a földrajzi egységek és a raktárak azok, amelyek nem sorolhatók a „megtörtént tranzakciók” körébe.

https://www.guru99.com/fact-table-vs-dimension-table.html

Vissza a kategóriáimhoz

Microsoft SQL (MS BI) (Training 360)
2. forduló

Ismertető a feladathoz

2. forduló

Tekintettel arra, hogy egy választ sem rögzítettél az alábbi feladatlapon, ebben a fordulóban a kitöltésére rendelkezésre álló idő teljes egésze, azaz 10 perc került rögzítésre mint megoldáshoz felhasznált idő.

1. feladat 0 / 5 pont

Banktól kapott szövegfájlt kell feldolgoznunk egy SSIS csomagban. A szövegfájlban szerepel a partner kódja, de ezt ellenőriznünk kell a dbo.Partner tábla adatai alapján. Milyen komponenst kell használnunk az ellenőrzéshez?

- ○ Merge
- ○ Script komponens
- ◉ Lookup
- ○ Union All
- ○ Union
- ○ Aggregate
- ○ Merge Join

Magyarázat a megoldáshoz

Minden bizonnyal ez a kérdés sem okozott komoly fejtörést. Sok oda nem illő doboz helyett nyilvánvalóan a Lookup transzformációs doboz az, amivel rákereshetünk egy érkező adatra egy meglévő adatbázis táblájában.

https://docs.microsoft.com/en-us/sql/integration-services/data-flow/transformations/lookup-transformation?view=sql-server-ver15

2. feladat 0 / 7 pont

A százas nagyságrendben elkészült csomagok tanulmányozásakor nyilván feltűnt Neked, hogy a leggyakrabban használt taszk a Data Flow taszk, és az ezekbe beépített transzformációk igen szerte ágazók. Felismered-e, hogy az alábbi felsorolásból melyek a transzformációs elemek?

- ☐ Data Transformation
- ☐ Excel Destination
- ☑ Merge Join
- ☐ Execute SQL
- ☑ DQS Cleansing
- ☑ Audit
- ☐ Update Statistics
- ☐ XML Source

Magyarázat a megoldáshoz

A válaszok között van pár gyanús szirén, amelyeknek nem szabad bedőlni. Data Transformation doboz például nincs, ez tehát csapda. Az Execute SQL és az Update Statistics nem Data Flow elemek, hanem Control Flow taszkok. Az XML Source és az Excel Destionation már Data Flow elemek, de nem transzformációs, hanem forrás, illetve cél elemek. Ha a tévutakat sikerült kizárni, akkor adódik a három helyes válasz: Merge Join, DQS Cleansing és Audit.

https://docs.microsoft.com/en-us/sql/integration-services/data-flow/transformations/integration-services-transformations?view=sql-server-ver15

3. feladat 0 / 5 pont

Egy SSAS projektben elvégzett szerkezeti módosításoknak be kell kerülniük a megfelelő multidimenzionális OLAP adatbázisba. Az adattárházban történt adatmódosításoknak is az OLAP adatbázisba kell kerülniük. Jelöld ki az alábbi kijelentések közül az igaz állításokat!

- ☐ A Deploy parancs a kocka adatait az adattárházba viszi fel
- ☑ A Deploy parancs a dimenzió adatait az Analysis Services adatbázisba viszi fel
- ☐ A Build parancs csak a dimenziók adatait viszi fel az OLAP adatbázisba, a Deploy veszi fel a mértékcsoportokat és a kockát
- ☑ A Process művelet az adattárházban történt adatmódosításokat dolgozza be az OLAP adatbázisba
- ☐ A Process művelet az OLTP adatbázisból viszi át az adatokat az adattárházba (DW-be)

Magyarázat a megoldáshoz

A kérdés a Build, a Deploy és a Process műveletek értelemzésére vonatkozott. A Build a megtervezett modell ellenőrzését és lefordítását végzi, a Deploy valósítja meg a modellt az OLAP adatbázisban, majd a Process az adatok besöprését (aktualizálását) hajtja végre. Az értelmetlen válaszok kiszűrése után a két jó megoldás marad.

https://ask.sqlservercentral.com/questions/121870/ssas-build-deploy-process.html

Vissza a kategóriáimhoz

Microsoft SQL (MS BI) (Training 360)
3. forduló

Ismertető a feladathoz

3. forduló

Tekintettel arra, hogy egy választ sem rögzítettél az alábbi feladatlapon, ebben a fordulóban a kitöltésére rendelkezésre álló idő teljes egésze, azaz 10 perc került rögzítésre mint megoldáshoz felhasznált idő.

1. feladat 0 / 5 pont

Az egyik riportban fel kell venned egy új Textbox elemet, és ennek a Dataset-ben található PartnerName nevű táblamező tartalmát kell megjelenítenie. Melyik kifejezés helyes az alábbiak közül?

- ○ PartnerName.Value
- ○ =Fields.PartnerName.value
- ○ Fields!PartnerName.Value
- ◉ =Fields!PartnerName.Value
- ○ =Fields!Partnername.Value
- ○ =Fields!PartnerName.value

Magyarázat a megoldáshoz

Ez a kérdés egyáltalán nem nehéz, de könnyű figyelmetlenül rossz helyre kattintani. Az SSRS alapvetően Visual Basic szintaktikájú, de – a VB-vel ellentétben – érzékeny a kisbetű/nagybetűre. Ebből adódóan a =Fields!PartnerName.Value a jó válasz.

2. feladat 0 / 8 pont

Az adatáramlási folyamatokat kell megtervezned, és döntened kell, milyen módon építed meg az ETL/ELT-t. Az alábbiak közül mely esetekben érdemes MS SQL Server Integration Services csomagokat tervezned az ETL/ELT-hez?

- ☐ Az adatáramlás forrása és célja ugyanazon az SQL Serveren található adatbázisban van
- ☑ SQL Server adatbázisból kell egy View alapján az adatokat szöveges fájlba exportálni
- ☐ Transact SQL tárolt eljárást kell futtatni az adatok áttöltéséhez
- ☑ OLAP adatbázist kell napi szinten frissíteni egy adattárház adatai alapján
- ☐ Tranzakcióban kell végrehajtanod több INSERT és UPDATE utasítást ugyanabban az adatbázisban

Magyarázat a megoldáshoz

Az ETL/ELT folyamatoknál mindig fontos szempont a hatékonyság. Ha egy részfolyamat megoldható az SQL Serveren belül (T-SQL scripttel), akkor az általában hatékonyabb megoldás, mint egy Data Flow Task. Mindezt figyelembe véve adódik a két jó válasz, amikor ki kell lépnünk a T-SQL világból.

3. feladat 0 / 5 pont

Elkészítettél egy SSIS csomagot, és Debug módban futtatod az SQL Server Data Tools-ban. Milyen ikon jelenik meg a taszk dobozának jobb felső sarkában a futás végén azoknál a taszkoknál, amelyekre nem került rá a vezérlés?

- ○ Piros „x” jel
- ○ Egyszerű fekete áthúzott kör
- ○ Sárga háromszög
- ◉ Nem jelenik meg semmilyen ikon
- ○ Zöld pipa
- ○ Lila kör
- ○ Fekete „x” jel

Magyarázat a megoldáshoz

Aki futtatott már Debug módban SSIS csomagot, az pontosan tudja, hogy semmilyen ikon sem jelenik meg azoknak a dobozoknak a sarkában, amelyek – az előfeltétel megszorítások következtében – nem kerültek végrehajtásra.

Vissza a kategóriáimhoz

Microsoft SQL (MS BI) (Training 360)
4. forduló

Ismertető a feladathoz

4. forduló

Tekintettel arra, hogy egy választ sem rögzítettél az alábbi feladatlapon, ebben a fordulóban a kitöltésére rendelkezésre álló idő teljes egésze, azaz 10 perc került rögzítésre mint megoldáshoz felhasznált idő.

1. feladat 0 / 5 pont

Kockát kell tervezned egy OLAP adatbázisban. Szépen összehordtad a szükséges dimenziókat, mértékcsoportokat és mértékeket, majd rámész az SQL Server Data Tools-ban a Dimension Usage kartonlapra. Mit definiál vajon itt egy cella?

- ○ Bármely két tetszőleges objektum (dimenzió, mértékcsoport, mérték) kapcsolódását egymáshoz
- ◉ Egy mértékcsoport és egy dimenzió kapcsolódását
- ○ Egy dimenzió és egy mérték kapcsolódását
- ○ Egy dimenziónak egy mértékcsoporttal vagy annak egy konkrét mértékével való kapcsolódását
- ○ Egy dimenzió attribútumnak egy másik dimenzió adott attribútumával való kapcsolódását

Magyarázat a megoldáshoz

A Dimension Usage kartonlapon kell összekapcsolni a dimenziókat a mértékcsoportokkal. Egy konkrét mértékkel nem lehet kapcsolatot létesíteni, csak egy komplett mértékcsoporttal.

https://docs.microsoft.com/en-us/sql/analysis-services/dimension-usage-cube-designer-analysis-services-multidimensional-data?view=sql-server-2014

2. feladat 0 / 6 pont

Az SSIS-ben egy Control Flow taszknál mit jelent a TransactionOptions=Supported beállítás?

- ○ Az adott taszktól kezdve a package végéig egyetlen tranzakcióban fut minden.
- ○ Az adott taszk mindenképpen tranzakcióban fut.
- ◉ Az adott taszk belép a felette lévő elem által létrehozott tranzakcióba, ha volt ahhoz tranzakció rendelve. Ha nem volt, akkor nem fut tranzakcióban.
- ○ Az adott taszk nem fut tranzakcióban.
- ○ Az adott taszk belép a felette lévő elem által létrehozott tranzakcióba, ha volt ahhoz tranzakció rendelve. Ha nem volt, akkor nyit egy új tranzakciót.

Magyarázat a megoldáshoz

Az SSIS-ben egy végrehajtható elem (taszk vagy konténer) esetén beállítható, hogy miként viselkedjen a tranzakció kapcsán. Az alapértelmezés a Supported, amire a kérdés vonatkozott. Ennél a beállításnál az adott végrehajtható elem belép egy létező tranzakcióba (ha van). Ha nincs nyitott tranzakció, akkor nem fut tranzakcióban. A Required nagyon hasonló a belépés tekintetében, de ha nincs létező tranzakció, akkor gyárt maga előtt egyet. A NotSupported esetén akkor sem lép be, ha amúgy korábban létre lett hozva.

https://www.mssqltips.com/sqlservertip/1585/how-to-use-transactions-in-sql-server-integration-services-ssis/

3. feladat 0 / 8 pont

Azt a feladatot kapod, hogy mérd fel, milyen előnyökkel és hátrányokkal járna, ha a cég adattárházát a felhőbe költöztetnétek. Mit tartalmazhat egy Azure Synapse Analytics adattábla az alábbiak közül?

- ☐ Clustered elsődleges kulcs
- ☐ xml adattípus
- ☑ bigint adattípus
- ☐ table adattípus
- ☑ IDENTITY
- ☐ Idegen kulcs

Magyarázat a megoldáshoz

Nem véletlen, hogy nem hemzsegnek a tesztkérdések között a felhős (Azure) kérdések. Ez a terület az, ahol rendkívül gyorsak a változások, így kockázatos feltenni egy kérdést, ami a tesztkérdések legyártása és adásba kerülése közötti 1-2 hónap alatt biztosan nem fog megváltozni. Még az adattípusoknál is lehet baj, mert például pár éve még nem létezett a varchar(max) a felhős környezetben. Az elsődleges kulcs, az idegen kulcs, az xml és a table adattípus remélhetőleg 2020-ban a nem létezők között fog maradni. Az alábbi link – kérdések megalkotásakor – 2020. június 1-ei állapotot mutatta:

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-data-types

Vissza a kategóriáimhoz

Microsoft SQL (MS BI) (Training 360)
5. forduló

Ismertető a feladathoz

5. forduló

Tekintettel arra, hogy egy választ sem rögzítettél az alábbi feladatlapon, ebben a fordulóban a kitöltésére rendelkezésre álló idő teljes egésze, azaz 10 perc került rögzítésre mint megoldáshoz felhasznált idő.

1. feladat 0 / 10 pont

A korábbi évek során számtalan riport készült SSRS-ben, és megkérnek rá, hogy az egyik riport fejrészébe helyezz fel egy szövegdobozt, ami az eredmény PDF-ben csak az első oldal tetején jelenik meg. Melyik szekcióba kell felraknod ezt a TextBox-ot?

- ○ Report Header
- ○ Page Header
- ◉ Body
- ○ Report Footer
- ○ Page Footer

Magyarázat a megoldáshoz

Mérgezett figurának hívják a sakkban azt a bábut, ami látszólag védtelenül lóg, és roppant csábító döntés leütni, ám a parti szempontjából végzetesek a következmények. Itt a Report Header válasz jelenti a mérgezett (csapda) választ, amit nem szabad bejelölni. Sok riporttervező eszköz – köztük az MS Access riporttervezője, ami az SQL Server Reporting Services őse – rendelkezik Report Header szekcióval, és valóban ide kellene helyezni a kért szövegdobozt. Ám az SSRS-nél nincs dedikált Report Header (se Report Footer). Itt a Body szekció az, amin a renderelés felülről lefelé végig halad, tehát ennek a tetejére kell helyezni bármit, amit csak egyszer, a riport elején szeretnénk látni az eredményben.

2. feladat 0 / 12 pont

A módosult forgalmi adatoknak a DW-be történő áttöltéséhez korábban Change Data Capture és Change Tracking technológiát használt a cég, de a 2016-os SQL Serverben megjelent Temporal (System-Versioned) táblák új lehetőségekkel kecsegtetnek. Melyek az igaz kijelentések az alábbiak közül?

- ☐ A System-Versioned táblában a SysStartTime és SysEndTime oszlopnevek kötelezőek, más mezőnév erre a célra nem használható
- ☑ A SysStartTime és SysEndTime mezők adattípusa kötelezően datetime2, sima datetime erre a célra nem használható
- ☐ A történeti (history) táblánál megadható idegen kulcs (foreign key constraint) megszorítás
- ☑ A System-Versioned táblánál kötelező elsődleges kulcsot (primary key constraint) definiálni.
- ☐ A System-Versioned tábla tetszőleges oszlopa módosítható UPDATE utasítással
- ☐ A történeti (history) táblában – a SysStartTime és SysEndTime kivételével – bármelyik mező módosítható UPDATE utasítással
- ☐ A tábla definiálásakor beállítható, hogy rendszeridőt vagy általunk megadott időt tárol a SysStartTime és SysEndTime mezőkben.

Magyarázat a megoldáshoz

A temporal tábláknál kell két dátum mező a táblában, de ezek neve szabadon választható, ám az adattípusuknak kötelezően datetime2-nek kell lennie. Ez a két mező nem módosítható UPDATE utasítással. A történeti táblában semmi sem módosítható. Elsődleges kulcs megadása kötelező. A tárolt dátum/idő csak rendszeridő lehet, ezért is hívják System-versioned-nek a szisztémát, és – sajnos egyelőre – nem létezik User-Versioned megoldás. Ez ugye akkor lenne jó, ha egy változás időpontjába nem a rögzítés idejét szeretnénk beramolni, hanem egy valamivel korábbi időpontot.

https://docs.microsoft.com/en-us/sql/relational-databases/tables/temporal-tables?view=sql-server-ver15

3. feladat 0 / 14 pont

Az OLAP adatbázishoz adatforrás nézetet (Data Source View) kell terveznünk. Az alábbi kijelentések közül jelöld meg az igaz állításokat!

- ☐ A megnevezett lekérdezéshez (named query) használt SELECT utasítás nem tartalmazhat HAVING klauzulát
- ☐ Az adatforrás nézetben SQL INSERT utasítással definiált virtuális tábla is szerepelhet
- ☐ Az adatforrás nézetben két tábla között csak olyan kapcsolat építhető fel, amelyik az adatforrás adatbázisában is létezett
- ☑ Az adatforrás nézetben új számított mező nemcsak a megnevezett lekérdezésként létrehozott virtuális táblához, hanem valós táblához is felvehető
- ☑ Az adatforrás nézetben bármelyik tábla bármelyik mezőjének a nevét átírhatjuk
- ☐ Egy Analysis Services projekt mindig csak egy adatforrás nézetet tartalmazhat
- ☑ Egy adatforrás nézet több különböző adatforrás tábláiból is épülhet

Magyarázat a megoldáshoz

Ez a kérdés az OLAP tervezéskor a Data Source View lehetőségeit járja körül. INSERT utasítással nem definiálható virtuális tábla, csak SELECT-tel, de ebben természetesen szerepelhet akár egy HAVING klauzula is. Olyan kapcsolat is megépíthető két tábla között, ami hiányzik a forrásul használt DW-ben. Több célirányos nézetet is létre tudunk hozni egy OLAP adatbázisban.

https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/defining-a-data-source-view-analysis-services?view=asallproducts-allversions

Vissza a kategóriáimhoz

Microsoft SQL (MS BI) (Training 360)
6. forduló

Ismertető a feladathoz

6. forduló

Tekintettel arra, hogy egy választ sem rögzítettél az alábbi feladatlapon, ebben a fordulóban a kitöltésére rendelkezésre álló idő teljes egésze, azaz 12 perc került rögzítésre mint megoldáshoz felhasznált idő.

1. feladat 0 / 7 pont

Két kirendeltségtől érkeznek szövegfájlok az ottani rendelésadatokkal. Mindkét kirendeltség egy központi rendelésszám osztó rendszert használ, tehát nincs két egyforma rendelésszám az érkező adatokban. Az érkező fájlok rendelésszámra rendezettek. Melyik dobozt kell használnunk, hogy a rendeléseket egy olyan – jelenleg üres – központi táblába írjuk be, ahol a Clustered elsődleges kulcs a rendelésszám?

- ○ Merge Join
- ◉ Merge
- ○ Multicast
- ○ Union
- ○ Union All
- ○ Copy Column

Magyarázat a megoldáshoz

Két, azonos rendezettségű szövegfájl összefuttatására a Merge transzformációs doboz használható abban az esetben, ha nem összekapcsolódó párokat kell keresnünk, hanem összefésülést kell végrehajtanunk. A párok keresésére a Merge Join lenne jó. A Union költségesebb megoldás, míg a Union All azért problémás, mert töredezettséget okozhat a Clustered indexben, hiszen nem tudhatjuk, hogy a két fájl milyen rendelésszám tartományt használ. A Multicast és a Copy Column jelenti a két, legkönnyebben kiiktatható hamis választ ennél a feladatnál.

https://docs.microsoft.com/en-us/sql/integration-services/data-flow/transformations/merge-transformation?view=sql-server-ver15

2. feladat 0 / 15 pont

A tételszintű eladási adatokat tartalmazó Fact.OrderDetail tábla mérete hamarosan eléri a 100 millió sort. A táblában 10-15 idegen kulcs, a degenerált dimenziókulcsként használható rendelés szám és tételszám mellett kb. 40 mérték szerepel. A táblába napközben többször is töltünk át adatokat a webshop adatbázisból. Az üzleti felhasználók által kiadott SELECT utasítások rendszerint 3-4 dimenzió mentén összegeznek 2-3 mértéket, például

```
SELECT CountryKey, SalesPersonKey, PromotionKey, SUM(Amount), SUM(Quantity)
FROM Fact.OrderDetail WHERE DueDateKey BETWEEN @FromDateKey And
@ToDateKey
GROUP BY CountryKey, SalesPersonKey, PromotionKey
```

Mivel a lekérdezések jellemzően egy adott időszakra vonatkoznak, particionált táblát javasolsz a DueDateKey mezőre építve.

Milyen index segíthet leginkább?

- ○ Clustered index a DueDateKey-re
- ○ Clustered index az idegen kulcsokra – a leggyakrabban használttól kezdve a ritkábbak felé
- ○ Nonclustered index az összes idegen kulcsra
- ○ Nonclustered index a leggyakrabban használt mezőkre fedőindexként
- ◉ Clustered Columnstore Index
- ○ Nonclustered Columnstore Index az összes dimenzió mezőre

Magyarázat a megoldáshoz

A DW-re végrehajtott csoportosító (GROUP BY) lekérdezések jellemzően végig olvassák az egész táblát, illetve megfelelően kialakított partícionálás mellett egy adott partíciót. Ezen semmilyen soralapú index nem tud segíteni – a fedőindexeket (covered index) leszámítva. Egy soralapú fedőindex viszont azért nem praktikus, mert gyakorlatilag a tábla összes oszlopának szerepelnie kellene benne. Marad tehát az oszlopos (columnstore) index, aminél az a nagy előny, hogy csak az adott lekérdezésben szereplő mezőket kell kihalászni, a többi mezőt nem kell végig olvasni. Ha csak a dimenzió mezőkre építünk indexet, akkor az nem jó, mert a mérték (measure) oszlopok nincsenek benne, és azok is kellenének.

https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-described?view=sql-server-2014

3. feladat 0 / 15 pont

Az üzleti elemzők panaszkodnak, hogy az adattárházban tárolt adatok pontatlanok, hiányosak, sok bennük a duplikátum. A főnököd – aki korábban egy szennyvíz-tisztító telepen dolgozott – lerajzolja Neked hogyan működik egy ilyen telep: egyik oldalon bejön a szennyvíz, a másik oldalon pedig a tisztítás után iható víz folyik a Dunába. Azt a feladatot kapod, hogy dolgozd ki annak a menetét, hogy a forrásrendszerekből érkező szennyezett adatokból tisztított adatok kerüljenek az adattárházba. Add meg az alábbi folyamatok helyes sorrendjét egy kilencjegyű szám formájában (pl. 571248369)

- MDS modell, entitások és attribútumok megtervezése, felépítése (1)
- MDS Excel Add-in révén Excelbe töltött adatokon duplikátum ellenőrzése, és a hibák javítása (2)
- SSIS csomag létrehozása DQS Cleansing transzformációval, ami a forrásrendszerből Staging táblába tölti az adatokat (3)
- MDS és DQS telepítése, konfigurálása (4)
- Domain adatok feltöltése (5)
- Matching policy definiálása (6)
- MDS entitás feltöltése Staging táblából (7)
- DQS tudásbázis (KB) felépítése, domain-ek létrehozása (8)
- MDS üzleti szabályok kialakítása (9)

A megoldások:
418956327
419856327

Magyarázat a megoldáshoz

Ez a kérdés egy sorba rendező feladat, ahol a Master Data Services (MDS) és a Data Quality Services (DQS) használatba vételének sorrendjét kellett jól kialakítani.

https://docs.microsoft.com/en-us/sql/master-data-services/master-data-services-overview-mds?view=sql-server-ver15

https://docs.microsoft.com/en-us/sql/data-quality-services/data-quality-services?view=sql-server-ver15

Vissza a kategóriáimhoz

Microsoft SQL (MS BI) (Training 360)
7. forduló

Ismertető a feladathoz

7. forduló

Tekintettel arra, hogy egy választ sem rögzítettél az alábbi feladatlapon, ebben a fordulóban a kitöltésére rendelkezésre álló idő teljes egésze, azaz 12 perc került rögzítésre mint megoldáshoz felhasznált idő.

1. feladat 0 / 16 pont

Azzal a feladattal bíznak meg, hogy készíts egy teszt adattárházat, amibe az éles adatok 1%-át kell átmásolnod véletlen kiválasztással. A Fact.Sales tábla kb. 100 millió sort tartalmaz, és pontosan 1.000.000 sort kellene átmásolnod. Melyik SELECT utasítás adja a leggyorsabb megoldást ehhez?

- ○ SELECT TOP 1000000 * FROM Fact.Sales
- ○ SELECT TOP 1000000 * FROM Fact.Sales ORDER BY RAND()
- ○ SELECT TOP 1000000 * FROM Fact.Sales ORDER BY RAND(SalesID)
- ○ SELECT TOP 1000000 * FROM Fact.Sales TABLESAMPLE(1000000 ROWS)
- ◉ SELECT TOP 1000000 * FROM Fact.Sales TABLESAMPLE(1200000 ROWS)
- ○ SELECT TOP 1000000 * FROM Fact.Sales ORDER BY NEWID()

Magyarázat a megoldáshoz

Ennél a tesztkérdésnél azt kellett mérlegelned, hogyan lehet pontosan 1 millió sort visszakapni véletlenszerűen egy 100 millió soros táblából a lehető leggyorsabban. A rendezettség nélküli TOP 1000000 a végrehajtási tervnek megfelelő legelső 1 millió sort adja vissza, ez tehát kicsit sem véletlenszerű. A RAND() függvénnyel ellátott rendezettség is ugyanezt adja, hiszen a paraméter nélküli RAND-ot csak egyszer hívja meg az SQL Server. A RAND(SalesID) vagy a NEWID() szerepeltetése az ORDER BY-ban már jó megoldás, ám érdemes megnézni a végrehajtási tervet, amiből kiolvasható, hogy ez rendkívül időigényes. A TABLESAMPLE sokkal gyorsabb, mert komplett Page-eket jelöl ki véletlenszerűen a táblából. Itt azonban az a baj, hogy ha pontosan 1000000 sort kérünk, akkor az sohasem lesz pont ennyi, hanem néha egy kicsivel több, néha kevesebb. Így aztán az a jó megoldás, ha egy kicsivel többet kérünk, és a TOP 1000000-val ennek a véletlen halmaznak az elejét kérjük.

https://www.mssqltips.com/sqlservertip/1308/retrieving-random-data-from-sql-server-with-tablesample/

2. feladat 0 / 12 pont

Néhány SSRS riporthoz gyorsítótár technológiát (Cached instance) szeretnénk igénybe venni. Mely kijelentések igazak az alábbiak közül?

- ☐ A gyorsítótárban tárolt jelentést csak az aktuális felhasználó az aktuális munkamenetben maximum 10 percig éri el.
- ☐ Egy adott jelentés csak egy példányban lehet jelen a gyorsítótárban.
- ☑ Egy adott jelentés – a lekérdezési paraméterkombinációtól függően – akár több példányban is jelen lehet a gyorsítótárban.
- ☐ A cached instance jelentéshez nincs szükség tárolt hitelesítő adatokra (stored credential), elegendő a futáskor megadni az adatforrásra vonatkozó bejelentkezési információkat.
- ☑ A cached instance jelentéshez minden esetben tárolt hitelesítő adatok szükségesek.
- ☑ Cached instance technológia nem használható az SQL Express verzióban.

Magyarázat a megoldáshoz

Cached Instance választása esetén minden paraméterkombinációhoz egy tárolt példány fog tartozni a megadott lejárati ideig, amit bármelyik felhasználó elérhet. Mivel a cache példánynál nem történik adatbázis elérés, ezért ez csak tárolt bizonyítvánnyal (stored credential) hozható létre. A Cache példány lejárati idejét az SQL Server Agent figyeli. Mivel az SQL Server Expressben nincs Agent, ezért ott nem tudunk a riportokhoz Cached Instance-t rendelni.

https://docs.microsoft.com/en-us/sql/reporting-services/report-server/caching-reports-ssrs?view=sql-server-ver15

3. feladat 0 / 17 pont

A CovidCsvImport nevű package adatokat olvas be egy csv fájlból a CovidCase táblába. A CovidCase táblába újonnan bekerült adatokat egy CovidCaseProcess tárolt eljárás dolgozza fel. Az egyik fejlesztő kolléga jobot javasol, aminek első lépése a CovidCsvImport package, a második lépés pedig a CovidCaseProcess tárolt eljárás meghívása. Egy másik kolléga az alábbi T-SQL scriptet javasolja – egyetlen jobstep-ként meghívni:

```
DECLARE @ExID bigint
EXEC SSISDB.catalog.create_execution
     @folder_name='Covid19',
     @project_name='Covid19Project',
     @package_name='CovidCsvImport.dtsx',
     @execution_id=@ExID OUTPUT
EXEC SSISDB.catalog.start_execution @ExID
EXEC CovidCaseProcess
```

Neked kell eldöntened, hogy a két megoldás egyformán helyes eredményt produkál-e.

- ○ A két megoldás ugyanazt az eredményt adja.
- ◉ Csak az első verzió (job két job step-pel) ad helyes eredményt.
- ○ Csak a második verzió (job egyetlen step-ként a fenti T-SQL scripttel) ad helyes eredményt.
- ○ Mindkét megoldás hiányos.

Magyarázat a megoldáshoz

Ez a kérdés azt járta körül, hogy mi a különbség egy package job-ból és T-SQL script-ből való meghívása között. A két megoldás között az a gyakorlati eltérés, hogy a T-SQl script-ből való meghívásnál a package külön szálon fut. A tárolt eljárás tehát nem várja meg a package futásának a végét. Ehhez be kellett volna rakni a script-be az alábbi utasítást:

EXEC SSISDB.catalog.set_execution_parameter_value @exID, 50, 'SYNCHRONIZED', 1

https://andyleonard.blog/2015/11/the-synchronized-ssis-execution-parameter/

Vissza a kategóriáimhoz