

CLOUD BI

6. forduló



A kategória támogatója: DXC Technology

Ismertető a feladatlaphoz

Kezdj neki minél hamarabb, mert a feladatot a forduló záró időpontjáig lehet beküldeni, nem addig lehet elkezdni!

Sok sikert!



Ebben a fordulóban elsősorban adattárházak működésével kapcsolatos feladatokkal találkozhatasz.

1. feladat 1 pont

Az alábbi állítások közül melyek jellemzik a **datalake**-et?

Válaszok

- ☐ jól meghatározott a felhasználás célja
- ☐ az adatforrásokból nyers formátumú adatok összessége
- ☐ viszonylag egyszerű új adatforrást implementálni, mivel nem szükséges komoly integráció az adattárház adatmodelljébe
- ☐ a tárolás és az elemzés együtt kerül megvalósításra
- ☐ általában strukturálatlan adat, illetve információ tömeget tartalmaz
- ☐ az integráció megvalósítása alacsonyabb költséggel jár
- ☐ támogatja az adatvezérelt innovációt a data scientist-ek, elemzők, üzleti felhasználók számára

2. feladat 1 pont

Mely állítás a legkevésbé igaz egy hagyományos, relációs adattárház fizikai megvalósítására?

Válasz

- ☐ ODS (Operational Data Store) az üzleti felhasználásra publikált köztes réteg, míg a STAGE a feldolgozás számára biztosított ideiglenes tárolási terület
- ☐ a surrogate key („pótkulcs”) segít az adat integrációban, független a forrásoktól az adat és típus változásoktól, időben állandónak tekinthető, általában generálódik és gyakran használják a dimenziók kapcsolásához
- ☐ a kiaknázás hatékonyabbá tétele érdekében sokszor denormalizáltan, redundánsan állítunk elő adatstruktúrákat
- ☐ a kiaknázások a magasabb szintű rétegekben (pl. adatpiacokban) gyakran használják a LOAD (érkeztetési) területre betöltött adatokat

3. feladat 1 pont

Hagyományos, relációs adatbázis környezetben nagy rekordszámú táblák töltésénél mely(ek) a legkevésbé alkalmazott irányelv(ek)?

Válaszok

- ☐ a nagy rekordszámú, „full” tábla töltését általában DELETE és UPDATE műveletekkel végezzük
- ☐ egy partícióban történő lekérdezés gyorsítására használható a lokális index

- ☐ kimagaslóan nagy rekordszámú táblák összekapcsolása esetén a megfelelő végrehajtási terv kiválasztásában elsődleges szempont az indexek megfelelő használata
- ☐ az adattárházak frissítési ciklusának megfelelően a lekérdezést a frissítési periódusnak megfelelő partícionálással segíthetjük

4. feladat 1 pont

A töltés típusokra vonatkozóan a következő állítások közül mely(ek) hamis(ak)?

Válaszok

- ☐ a tranzakciós adatok töltésénél (pl. bankban a folyószámla napi tranzakcióinak tárolásakor) mindig SCD2 töltési stratégiát használunk
- ☐ az SCD1 („full”) töltésű tábla az összes adat aktuális állapotát tartalmazza
- ☐ az SCD2 tábla szervezésének alapja valamilyen érvényességi periódus
- ☐ a tényadatok általában nem kapcsolódnak időhöz, azaz sokszor nincs időbeliségük
- ☐ az SCD2 töltéshez kapcsolódóan gyakran használunk ún. history táblát az előző érvényességű állapotok tarolására, így például az aktuális állapotú adatok lekérdezése gyorsítható
- ☐ az érvényességi időszak mindig alulról zárt, felülről nyitott intervallumként értelmezendő

5. feladat 1 pont

Melyik állítás(ok) nem igaz(ak) a **dimenzionális modellekre**?

Válaszok

- ☐ a dimenziók nem tartalmazhatnak numerikus adatot
- ☐ tárolhatunk olyan dimenziókat, melyek nem zárt rendszeren belüli folyamattal állítódnak elő, hanem külső, kézíleg szerkesztett adatforrásból (pl. csv, excel, txt, xlm)
- ☐ a tényadatokon kialakított hierarchiák lehetővé teszik az adatok részletes vagy összefoglaló elemzéseit
- ☐ lefűrás csak a mérőszámok mentén lehetséges
- ☐ adott ténytáblához kapcsolódhat egy vagy több dimenzió

6. feladat 1 pont

Mely állítás(ok) jellemző(ek) a felhőalapú (cloud) adattárházakra?

Válaszok

- ☐ Elasztikus teljesítmény
- ☐ Csekély külső hálózati forgalom
- ☐ Teljes szabadsággal megvalósítható egyedi integrációk
- ☐ Helyfüggetlen, rugalmas hozzáférés
- ☐ Költséghatékonyság
- ☐ Beruházás megtérülése csak hosszú távon várható

7. feladat 1 pont

Az alábbi személyek esetében, ki(k)re a legkevésbé igaz a következő állítás: **ők az adattárház legismertebb szakértői?**

Válaszok

- ☐ Larry Ellison
- ☐ Ralph Kimball
- ☐ Edgar F. Codd
- ☐ Raymond F. Boyce
- ☐ Claudia Imhoff
- ☐ Bill Inmon

8. feladat 1 pont

Melyik a kakukktojás az alábbiak közül?

Válasz

- ☐ AWS Glue

- ☐ Azure Data Factory
- ☐ Google Cloud Dataflow
- ☐ Exadata
- ☐ SnapLogic
- ☐ Oracle Data Integrator Cloud Service (ODICS)

Megoldások beküldése