

Név:

Minden helyes válasz egy pont. De ha egy kérdésen belül rossz választ is ad, az kioltja a jó válaszáért kapott pontot. Ha nem volt jó válasz a kérdésre, 0 pontot kap.

Ha több jó helyes megoldás is van, az látszik a feltett kérdés után.

Előfordulhat, hogy egyes esetekben a kérdésre adott válasz átkerült a következő oldalra.

1. Mikor indul egy Spark Streaming alkalmazás?

- a. Amikor az olvasási adatfolyam létrejön.
- b. Amikor az átalakítási folyamat létrejön.
- c. Amikor az írási adatfolyam létrejön.
- d. A start() metódus meghívásakor.

2. Melyik jellemző leginkább a MongoDB adatbázisokra?

- a. Kulcs-érték tároló.
- b. Oszloptároló.
- c. Dokumentum tároló.
- d. Sortároló.

3. Egy streaming DataFrame-en végrehajtott művelet a következőt eredményezi:

- a. Streaming DataFrame.
- b. Statikus DataFrame.
- c. A művelettől függ mit kapunk.
- d. Nem tudunk műveleteket végrehajtani a streaming DataFrame-en.

4. Melyik tulajdonság jellemző a Cassandra adatbázisra? (Több helyes válasz is jó lehet?)

- a. Bár flexibilis szerkezetű, de a felhasználóhoz kapcsolódó felülete SQL alapú.
- b. Az oszlopok nevei soronként változhatnak.
- c. Nem köti az oszlopok névsor alapú rendezése sem.

5. Melyik tulajdonság jellemző a Spark adatbázisokra?

- a. Az adatokat a cach memóriában tárolja, így gyors számítások végezhetők rajta.
- b. Az adatokat dokumentumokban tárolja, melyben beépített tömbök és dokumentumok is előfordulhatnak.
- c. Az oszlop nevek adják az adatállományt.

6. Meddig tárolja az Apache Kafka az üzeneteket?

- a. Amíg az első felhasználó el nem olvassa.
- b. Amíg az összes üzenetet el nem olvasták.

- c. Amíg a megőrzési/visszatartási idő le nem telik.
- d. Amíg kifejezetten/véglegesen nem törlik.

7. A Spark strukturált adatfolyam az adatfolyamokat úgy kezeli, mint

- a. egy táblázatot az utolsó feldolgozatlan sorokkal.
- b. egy korlátlan táblát.
- c. egy korlátos táblát.
- d. egy üzenetsort.

8. Mik az előnyei a No SQL adatbázisoknak?

(Több helyes válasz is jó lehet.)

- a. Könnyen átlátható, egyszerűen kezelhető struktúra.
- b. Flexibilis szerkezet.
- c. Könnyen skálázható.
- d. Egy helyen tárolható minden adat.

9. Milyen zárolási problémák okozhatnak gondot a hagyományos relációs adatbázisoknál?

- a. Soha nem kapcsolható ki a szerver.
- b. Ha egyszer zároltuk a szervert nem tudni, mikor érhető el újra.
- c. Holtpontok keletkezhetnek.

10. Milyen paranccsal indítható Cassandra-ban egy cluster?

- a. Use cluster
- b. open cluster
- c. start cluster

11. Az alábbi összekapcsolási (join) műveletek közül melyik fordulhat elő a Sparkban?

(Több helyes válasz is jó lehet.)

- a. Összekapcsolódik két statikus DataFrame.
- b. Összekapcsolódik két streaming DataFrame.
- c. Összekapcsolódik egy statikus és egy streaming DataFrame.

12. Mikor ér véget egy Spark Streaming alkalmazás?

- a. Ha nem érkezik üzenet a Vízjel (Watermark) által meghatározott ideig.
- b. Amikor feldolgozza az adatfolyam összes üzenetét.
- c. Ha véglegesen megszűnik.
- d. Amikor a bemeneti adatfolyam leáll.

13. Mik tartoznak a relációs adatbázis előnyei közé?

- a. Bárki tudja kezelni, bármilyen előismeret nélkül.
- b. Jól bevált, jól ismert, általános felhasználási célú, megbízható.
- c. Könnyen skálázható, olcsón fejleszthető.

14. Mit jelent a horizontális skálázhatóság?

- a. A szerver adattárolási és kiolvasási gyorsaságának növelése.
- b. Több párhuzamos szerver bevonása.
- c. A memória egyre nagyobb léptékekkel való bővítése.

15. Melyik jellemző leginkább a Cassandra adatbázisokra?

- a. Kulcs-érték tároló.
- b. Oszloptároló.
- c. Dokumentumtároló.
- d. Sortároló.

16. Kell-e elsődleges kulcsot megadnunk a Cassandra tábláknál, és ha igen mikor?

- a. Rögtön, amikor a táblát létrehozuk.
- b. Nem muszáj.
- c. Elegendő a meglévő táblához utólagosan hozzáadni egy megkötést (constraint).

17. Melyik hierarchikus (egymásba ágyazódó) adattárolási rendszer figyelhető meg a MongoDB használatakor?

- a. Adatbázis / gyűjtemények / dokumentumok / kulcsok
- b. Adatbázis / táblák / rekordok / mezők

18. Melyik állítás helyes a MongoDB-re vonatkozóan? (Több helyes válasz is jó lehet.)

- a. Ez egy dokumentum-orientált adatbázis.
- b. Ez egy erre a célra épített tárolórendszer, amelyet kifejezetten nehéz írási igényekre terveztek.
- c. Master nélküli architektúrája van.
- d. Elosztást/feldarabolást (sharding) használ a vízszintes skálázhatóság érdekében.

19. Mit nem lehet megtenni a MongoDB-ben?

- a. Tárolni a különböző szerkezetű dokumentumokat egyetlen gyűjteményben.
- b. Végrehajtani olyan lekérdezést, amely több elosztásból (shard) is lekéri az eredményeket.
- c. Eldobni az elsődleges kulcs mezőt (_id) a dokumentumból.
- d. Frissítési nyilatkozatot tenni azokról a dokumentumokról, amelyek megváltoztatják a szerkezetüket.

20. Melyek a helyes használati lehetőségek arra, hogy a kliensek csatlakozzanak egy replikakészlet másodlagos példányaihoz?

- a. A rendszer írási sebességének növelése.
- b. A válaszok következetességének javítása.
- c. Kiegyensúlyozottabb adatelosztás engedélyezése a cluster példányai között.
- d. Magas rendelkezésre állású olvasási lehetőség biztosítása még elsődleges hiba esetén is.

21. Mire vonatkozik a „No SQL” megnevezés?

- a. Egyáltalán nem használ SQL-t. (Not Organized by SQL.)
- b. Nem csupán SQL-re támaszkodik. (Not Only SQL.)
- c. Tiltakozás az SQL ellen. (Never use mOre SQL.)

22. Az alábbi állítások közül melyik igaz az elosztással (sharding) kapcsolatban?

- a. Az elosztott rendszerrel több adatot tárolhat, mint amennyi egyetlen gépen elérne.
- b. A kliens közvetlenül kommunikál a megosztott példányokkal.
- c. Az elosztott kulcsmező szűrése jelentősen javíthatja a lekérdezés teljesítményét.
- d. Az adminisztrátornak rendszeresen át kell helyeznie az adattömböket az elosztott példányok között, hogy megbizonyosodjanak arról, hogy az adatelosztás nem válik kiegyensúlyozatlanná.

23. Mely elveket, kell követni a MongoDB adatmodelljének tervezésekor?

(Több helyes válasz is jó lehet.)

- a. Minden entitást a saját gyűjteményébe kell helyezni, hogy az adatok normalizálódjanak.
- b. Általában jó ötlet a módosítandó mezők együttes tárolása ugyanabban a dokumentumban.
- c. Indexeket használni a gyakran szűrt vagy rendezett mezőkre.
- d. Kevés számú elosztási kulcs kiválasztása, hogy elkerüljük a túl sok kis adatdarab létrehozását.

24. Mely állítások igazak a Big Data technológiákkal kapcsolatban?

- a. Céljuk a nagy adatmennyiség, sebesség és változatosság problémáinak megoldása, ami a hagyományos adattechnológiákkal nem lehetséges.
- b. Operatív adatbázisként szolgálnak nagyobb adatkészleteken, amelyek az egyes rekordok kezelésére összpontosítanak.
- c. A legtöbb esetben valós idejű válaszokat várunk a Big Data rendszerektől ezredmásodpercek sorrendjében.
- d. Az adatlokalitás fő gondolata az adatok másolása az algoritmust futtató számítógépre.

25. Mely állítások igazak a Sparkra?

- a. Ez egy adatfeldolgozó motor, amely elosztó módon működik nagy mennyiségű adaton.
- b. Az összes közbenső számítási eredményt lemezen tárolja, hogy hibátűrést biztosítson.
- c. Jó jelölt egy működő adatbázishoz, mivel rendelkezik teljesítő OLTP-képességekkel.
- d. Modulokat tartalmaz, amelyek támogatják a gépi tanulást, az adatfolyam-feldolgozást és a gráfszámításokat.

26. Mely állítások igazak a Spark műveletekkel kapcsolatban? (Több helyes válasz is jó lehet.)

- a. Minden transzformáció azonnal megvalósítja eredményét.
- b. Az átalakítás eredménye mindig egy új adatgyűjtés (pl. RDD, DataFrame).
- c. A keskeny átalakítások általában alacsony költségű műveletek a széles átalakításokhoz képest. IGAZ
- d. Példák az átalakításokra a szűrők, vetítések és az összekapcsolások is.

27. Az alábbi API-k közül melyik biztosít a fordítási idő biztonságát?

(Több helyes válasz is jó lehet.)

- a. RDD.
- b. DataFrame.
- c. Adatkészlet.
- d. SQL lekérdezés.

28. Létrejön-e kulcssértés akkor, ha a Cassandra-ba ugyanazzal a kulccsal másik értéket akarunk bevinni.

- a. Mindenképpen.
- b. Csak akkor, ha az insert parancsot használjuk.
- c. Nem. Egyszerűen felülírja a korábbi értéket.

29. Mit tesz a Cassandra, ha olyan rekordot akarunk az update paranccsal módosítani, amelyik nem is létezik?

- a. Hibajelzéssel tér vissza.
- b. Létrehoz egy új sort a megadott paraméterekkel.
- c. Jelzi, hogy olyan feltételt adtunk meg, ami nem egyezik az adatbázis adataival.

30. Mely állítások igazak a Hadoopról? (Több helyes válasz is jó lehet.)

- a. Míg eredetileg csak maroknyi egymással együttműködő komponenst tartalmazott, azóta különféle technológiák hatalmas ökoszisztémájává vált.
- b. A Map-Reduce keretet biztosít az elosztott adattároláshoz.
- c. A HDFS nagyszámú kis fájl tárolására van optimalizálva.
- d. A Hive maga nem adatbázis, hanem SQL-szerű hozzáférést biztosít az elosztott tárolórendszeren (pl. HDFS) tárolt adatokhoz.