



Az adattárházról egyszerűen – pillantás a fekete dobozba?

Az adattárház az Oracle kínálatában nem fekete doboz. Nagyon is transzparens, hiszen csakis általánosan elterjedt és szabványos kapcsolódási felületekkel rendelkező alkatrészekből áll.

■ Az adattárház definíciója szerint adatokat gyűjt, tárol és kínál fel másodlagos (tehát nem elsősorban tranzakciós típusú) hasznosításra. Több témakör adatait felöleli – a vállalati adattárház alapvetően az összes elemzendő témakörből –, és teszi elérhetővé elsősorban jelentésekben, lekérdező és elemző eszközökön keresztül. A másodlagos hasznosítás itt nem az adattárház környezetvédelmi szerepére utal, hanem arra, hogy az alaptevékenység adminisztrálásán (pl. számlavezetés, számlázás, események regisztrálása, stb.) túlmenően a vállalatok és szervezetek által gyűjtött adatokból egy „mélyebb értelmezéssel”, elemzéssel a vállalat, vagy intézmény működési hatékonyságának, az ügyfelek viselkedésének jellemzőit tudjuk szintetizálni; majd üzleti intelligencia eszközökkel megjeleníteni.

Ezen elemzési igények kiszolgálása az informatika fejlődési irányait elemző cégek szinte egyöntetű véleménye szerint elsődleges prioritássá vált az utóbbi évben (pl. *2007 Gartner EXP – CIO Survey, 2006 Merrill Lynch CIO Survey*).

Ne titkoljuk el, hogy sok jelentési, elemzési feladatot meg lehet oldani adattárház nélkül is, hiszen az Oracle kínálatában ma már léteznek olyan üzleti intelligencia eszközök, amelyek közvetlenül az operatív adatbázisok adatain is képesek akár federatív módon, több tématerület feldolgozását támogató, adatbázisokból egyidejűleg kombináltan is lekérdezéseket végrehajtani, akár heterogén környezetben

is. Vannak azonban esetek, amikor nem tudjuk elkerülni az adattárház építését. Néhány példa ezekre az esetekre:

- az adatok nem őrződnek meg kellően hosszú ideig a tranzakciós rendszerek adatbázisaiban
- az elemző lekérdezések akadályozzák, lassítják a tranzakciós munkát
- az alaprendszeri adatséma túl bonyolult a lekérdező eszközök számára
- az egyszerű aggregációkon és kalkulációkon túlmenően bonyolult algoritmusokat is be akarunk vetni új információ szintetizálására
- jelentős adatminőség javításra van szükség
- biztonsági okokból nem akarunk hozzáférést adni egy csoport számára egy, vagy több alaprendszeri adatbázishoz

Amennyiben csak egy témakörrel foglalkozik nevezhetjük adatpiacnak; jelentés adatbázisnak – be elsősorban jelentések készítését támogatja (mi más tehetne); vagy integrációs platformnak, ha az adatintegrációs képességét akarjuk kihangsúlyozni; vagy használhatjuk szívünk kedvenc szavát is, amennyiben az adattárház szóhoz a kiszemelt belső szponzorok azonnal horribilis költségeket asszociálnának. Az adattárház építés kétségtelenül az informatika egyik legizgalmasabb, nagy áttekintést ígérő és ugyanakkor legkockázatosabb vállalkozása. Ha fenti pontok alapján adattárház építésbe kell kezdenünk, akkor pillant-

sunk bele dobozunkba. Érdeemes, hiszen, az adattárház intenzív kapcsolatban van a környezetével. Egy kiterjedt adattárháznak rengeteg ki és bemeneti ponton van kapcsolata az alap informatikai rendszerekkel, a felhasználókkal, fejlesztőkkel és üzemeltetőkkel. Ilyen értelemben nem érdemes, és nem is lehet, fekete dobozként kezelni. Az adattárház belsejét érdemes elterjedt, és szabványos alkatrészekkel kitölteni:

- egyrészt azért, mert a *TPC-H benchmarkok* (lásd. *tpc.org*) tanúsága szerint az elterjedten használt általános célú relációs adatbázis kezelők jobban szerepelnek teljesítmény és ár/érték jellemzők alapján is, mint a specializált adatbázis motorok
- másrészt mert a hosszú távú költségszámításokban (*Total Cost of Ownership – TCO*) már nem elhanyagolható, hogy az egyébként is elterjedten használt adatbázis kezelőkhöz házon belül, és a piacon kiterjedten férhető hozzá a tudás, amivel a rendszer életben tartása, és továbbfejlesztése megoldható.
- szintén nem elhanyagolható erejű érv az, hogy az általános célú adatbázis-kezelő security, hordozhatóság és skálázódás szempontból jelenleg éveket ver rá specializált társaira.

Linuxon az Oracle a következő TPC-H rekordokat tartja:

- 300 gigabyte
- 3 terabyte

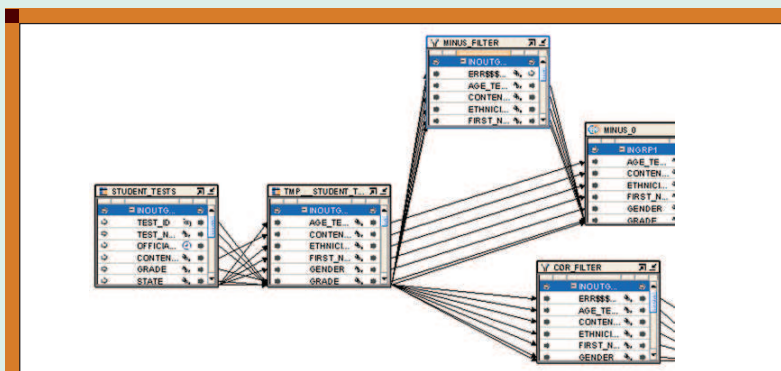


A *Winter Corporation* által auditált rendszerek között a világ legnagyobb *Linuxon* futó adattárháza az *Amazon* által működtetett több mint 24 terabyte-os *Oracle* alapú adattárház. Az egyéb (nem adattárház és nem *OLTP*) kategóriában is az *Oracle* viszi el a pálmát a *Max Planck intézet* *Linuxon* futó 222 terabyte-os adatbázisával.

Az *Oracle* adatbázis-kezelője több piacutató cég szerint egybehangzón 40%-át fedi le a relációs adatbázis alapon épített adattárházak piacának (pl. *IDC* 2006 adattárház piac).

Az *Oracle* adatbázis-kezelője a különböző cég méretekre és adatbázis-kezelési igényekre alapozva többféle változatban is kapható. Különböző méretű adattárházi, vagy adatpiaci projekteken sikerrel alkalmazzák az ingyenes legkisebb adatbázis kezelőtől az *Enterprise* változatig bezárólag mind egyik verzióját. Ugyanakkor mind egyik *Oracle* adatbázis-kezelő ugyanazt a kódot futtatja, melyek között az áttérési lehetőséget a cég egyszerűen biztosítja, amellett, hogy az egyik változatba fektetett munka nem vész el az áttéréskor.

Természetesen az *Oracle Database* és az *Oracle* üzleti intelligencia eszközök is számos platformon, így *Linuxon* is működnek. Ha valaki például szeretne áttérni *Linux* platformra, roppant egyszerűen migrálhat *Linuxra*. Újdonság, hogy az adatfeltöltésre – és a megfelelő adatelrendezést biztosító *Oracle Warehouse Builder* eszköz ingyenes kiegészítőként bekerült az *Oracle* adatbázis-kezelő minden megvásárolható kiadásába. Bonyolultabb esetekre, amikor a fejlett adattisztítás,



■ 1. ábra Megfelelő adatelrendezés az Oracle Warehouse Builder segítségével

vagy a több környezetben történő futtatás is igény, az ingyenes alap mellé kiterjedt funkcionalitással rendelkező fizetős opciókat is kínál a cég, csakúgy mint az *OLAP* és adatbányászati és particionálási funkciók támogatására. A *Warehouse Builder Data Correction* varázslója automatikusan generál adat javító függvényeket. Az alábbi példában az *SDX_CONTENT_AREA* függvény látható.

```

type t_arr is table of
↳ varchar2(2000) index by
↳ binary_integer;
s t_arr;
begin
  s(0) := 'MATH'; s(1) :=
↳ 'READING'; s(2) :=
↳ 'SCIENCE'; s(3) := 'WRITING';

  - check for equality first.
↳ Function may be called when
↳ there is no error.
  for i in s.first..s.last loop
    if (s(i) = CONTENT_AREA)
↳ then
      return CONTENT_AREA;
    end if;
  end loop;

  for i in s.first..s.last loop
    if (soundex(s(i)) =
↳ soundex(CONTENT_AREA)) then
      return s(i);
    end if;
  end loop;
  return CONTENT_AREA;
end;
```

A függvény lefuttatását követően például az angolul hibásan írt *WRITEING* automatikusan javításra kerül a helyes *WRITING* szóra.

Az *Oracle* alapú adattárházban gyűjtött adatok kiaknázását számos gyártó lekérdező és elemző eszközkészlete támogatja. Érdemes mégis megismerni, és kipróbálni az *Oracle* kínálatában teljesen megújult üzleti intelligencia csomagokat, melyek egy dashboard keretben biztosítják a jelentéskészítést és -megtekintést, és az ad-hoc lekérdezések és a különböző eszközökre küldhető figyelmeztetések kiküldését. A jól megválasztott szoftver eszközpark persze csak egy összetevője a sikeres adattárház projektnek. Mivel a kockázatok csökkentésében jelentős szerepe van a szabványos elemek használatának, ne csak nézzünk bele a dobozba, hanem válogassunk addig a számunkra szükséges alkatrészekből, amíg a problémának megfelelő rendszert sikerül összerakni.



Radnai Szabolcs

(szabolcs.radnai@oracle.com)

Radnai Szabolcs az *Oracle Hungary Kft.* 1993-as megalakulása

óta dolgozik az *Oracle*-nél. Tíz éve foglalkozik üzleti intelligencia és adattárház megoldási javaslatok kidolgozásával. Jelenleg az Üzleti Intelligencia és Adattárház üzletág vezetője.

KAPCSOLÓDÓ CÍMEK

- ↳ <http://oracle.com/bi>
- ↳ <http://www.gartner.com/it/page.jsp?id=501189>
- ↳ http://www.gartner.com/press_releases/asset_152619_11.html