



Memóriarezidens adatbáziskezelés

Marton József Ernő
marton@db.bme.hu



Adatbázisok haladóknak 2012.

2012. szeptember 18.



Miről lesz szó?

Memóriaadatbázisokról általában

Alapelv, kontraszt a „hagyományossal”

Előnyök, erősségek

Hátrányok (?), nehézségek

Szervezés

Kommunikáció

ACID *D* betűje

Fizikai ~, Lekérdezésvégrehajtás

Megvalósítások

Kérdések, válaszok

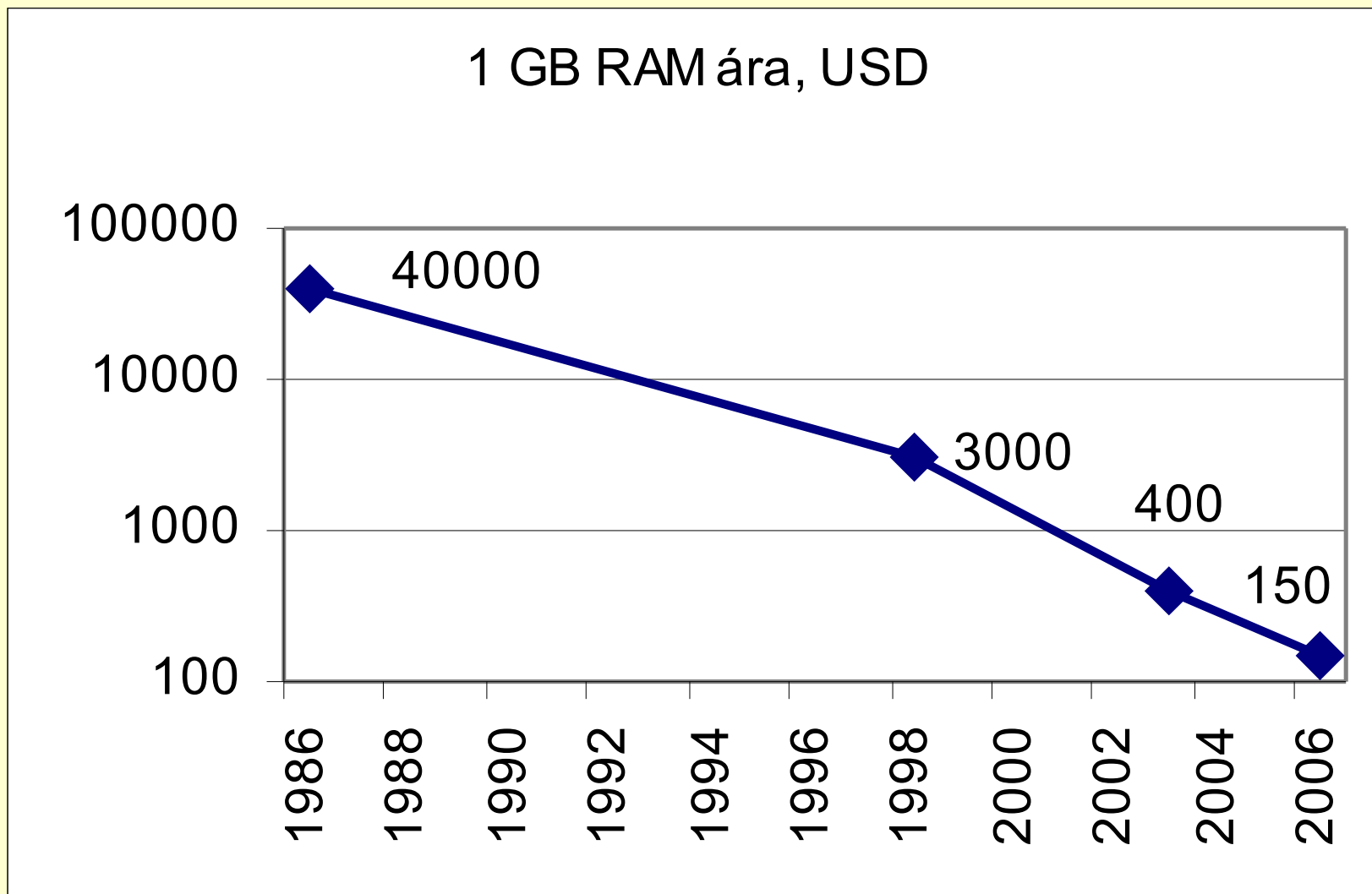


Memóriaadatbázis: alapelv

- Memóriaadatbázis-rendszer (IMDB)
(In-Memory DataBase)
 - az adatok elsődleges példánya a fizikai memóriában
 - DRDB: az elsődleges példány diszken van
(Disk-Resident DataBase)
 - mondjuk inkább úgy: blokkos tárolón (vö. SSD)



Az alapelv létjogosultsága (2012: 20 USD)





Az alapelv létjogosultsága (2)

- Sok és olcsó RAM
 - adatbázisok, adattárházak
- Még nagyobb igény esetén: adatbázis particionálása IMDB és DRDB részre
 - automatikusan vagy kézzel konfigurált módon
 - az adatok migrációja a két rendszer között



IMDB használatának előnyei

- gyorsabb adatelérés, tranzakció-feldolgozás
 - akár hard real-time alkalmazások
- egyes esetekben egyszerűbb alkalmazás-logika
 - diszk-IO alrendszer eliminálása
 - $\Rightarrow$ még gyorsabb működés
- saját adatszerkezetek helyett kész komponens
 - olcsóbb fejlesztés, olcsóbb termék
- a termék sikerét is meghatározhatja



IMDB: hátrányok, nehézségek

- bizonyos esetekben bonyolultabb alkalmazás-logika
 - tartósság biztosítása
 - biztonsági mentések szervezése
- a rendszer indulásakor fel kell tölteni az adatbázist
 - persze leállítás után is
 - mentésből, fejállomásról letöltve vagy on-line adatgyűjtésből
- Megéri-e: előnyök vs. hátrányok
 - a konkrét feladat határozza meg



IMDB: megvalósítási kihívások

- optimalizált adatszerkezetek
 - a fizikai memória „random” elérését kihasználják
- zárkezelés, tranzakciók ütemezése
 - megfelelő granularitású zárok $\Rightarrow$ egyszerűbb alkalmazáslogika
- trükkök a tartósságra
 - HW támogatás: elemes RAM, hálózaton szinkronizáló RAM-kártyák
 - szinkron vagy aszinkron naplózás
 - nem követelmény ☺

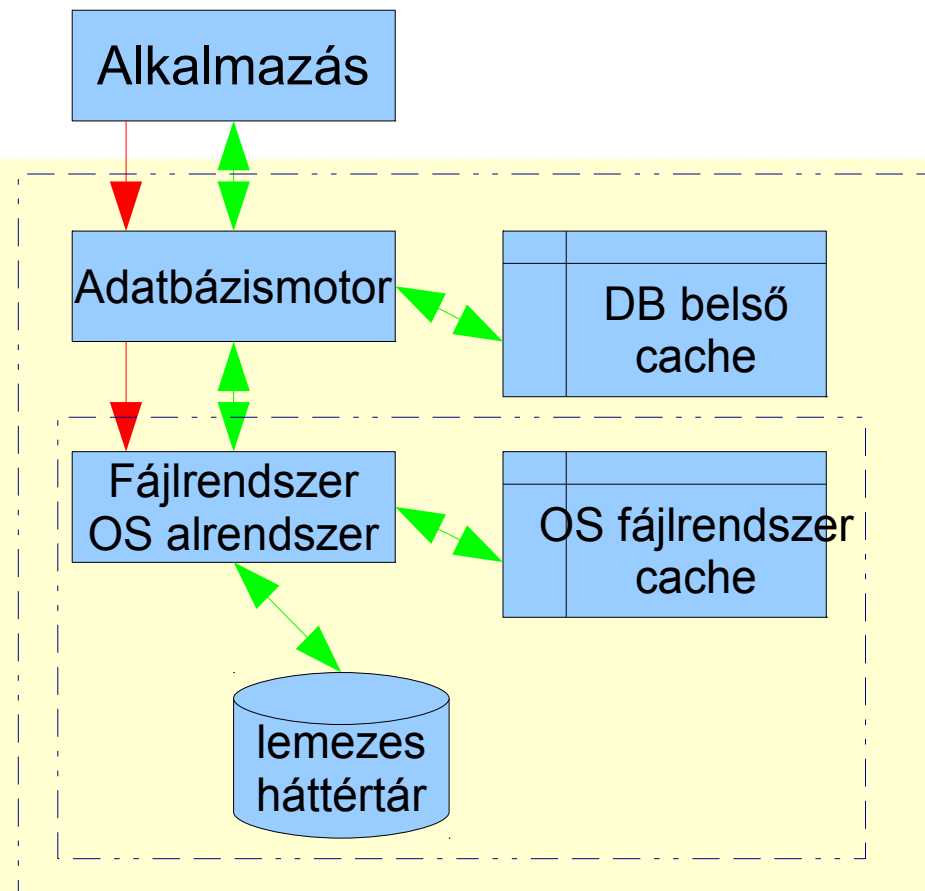


Vajon IMDB-nek minősül?

- egyedi alkalmazás saját adatszerkezetekkel
- hagyományos DBMS úgy, hogy
 - nagy fájlrendszer cache az operációs rendszerben
 - adatfájlok RAM-diszken
 - nagy a buffer cache



Kommunikáció

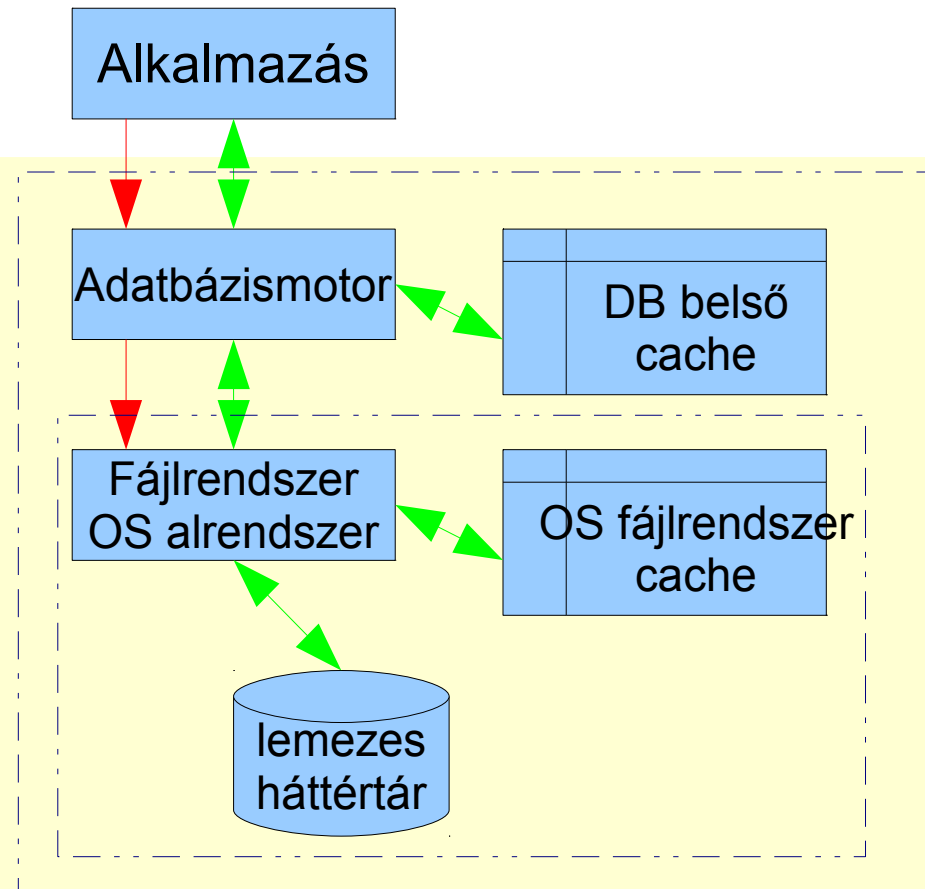


*Adathozzáférés komponensei DRDB esetén
a piros nyilak üzenetátadást, míg a zöldek
adatáramlást jelölnek
a szaggatott keret átlépése a futási
környezetváltásokat mutatja*



Kommunikációs alternatíva

- kliens-szerver modell
- osztott memória alapú adatelérés
 - durva hozzáférés szabályozás



*Adathozzáférés komponensei DRDB esetén
a piros nyilak üzenetátadást, míg a zöldek
adatáramlást jelölnek
a szaggatott keret átlépése a futási
környezetváltásokat mutatja*



Tranzakciós tulajdonságok

- A – atomicitás
 - nincs markáns különbség
 - gyorsabb tranzakciók, magasabb zár-granularitás
- C – konzisztencia
 - nincs markáns különbség
- I – izoláció
 - sorosítható helyett valódi soros ütemezés (vö. processzor-cache, mint környezet)
- D – tartósság



D – tartósság

- Az adatbázisba „írt” adatok „megmaradnak”.
- diszk: passzív
- memória: aktív
 - tápfeszültség kimaradásakor törlődik
 - elemmel támogatott RAM
- bithiba – modul meghibásodások
 - vö: diszkek: RAID
 - modultöbbszörözés
- naplózás (és persze mentések)



D – tartósság: naplózás

- ... de hova?
 - diszkre: már megint diszkIO
 - stable memory
 - kicsi, nagy megbízhatóságú memória-megoldás
- ... de mikor?
 - írási műveletkor azonnal
 - tranzakció commitjakor
 - még később?!
 - ún. durable commit (Oracle TimesTen)



Táblázat: durable commit hatása a tranzakciófeldolgozási kapacitásra

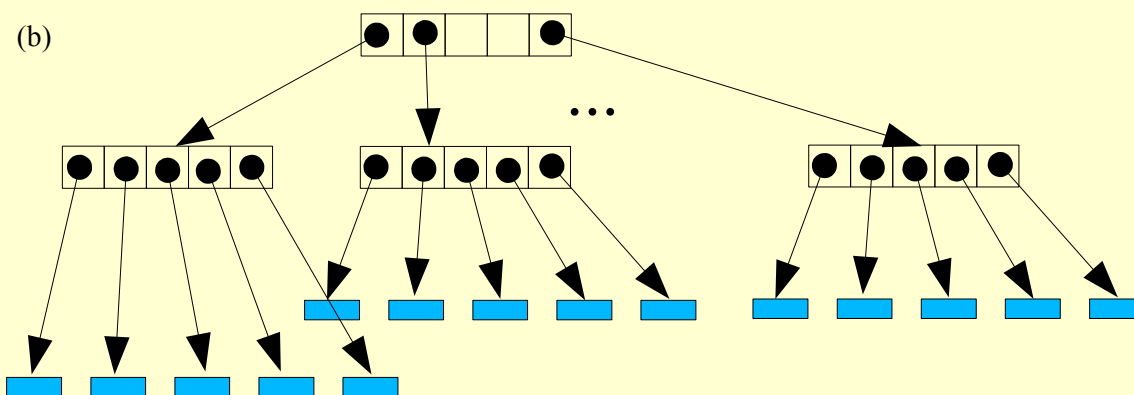
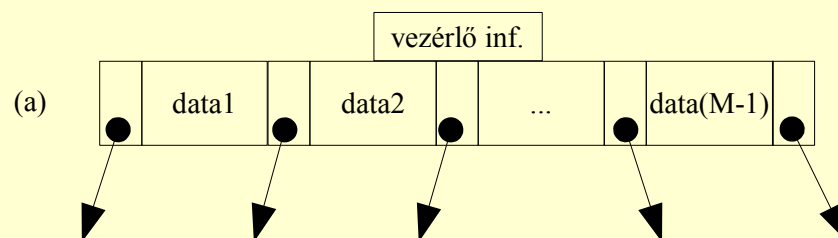
<i>select/insert/update</i>	<i>70/15/15</i>	<i>40/30/30</i>	<i>20/40/40</i>
<i>tartós commit</i>	<i>100 %</i>	<i>52 %</i>	<i>38 %</i>
<i>nem tartós commit</i>	<i>918 %</i>	<i>714 %</i>	<i>626 %</i>

- AMD Athlon64 3000+ (1809 MHz) processzorral és 4 GiB fizikai memóriával rendelkező gép teljesítménye egy 7200 rpm fordulátú Seagate Barracuda UltraATA 100-as merevlemezrel és a Debian Linux 3.1 operációs rendszer AMD64-es változatával, amelyen a TimesTen 6.0.2-es, 64 bites verziója fut(ott)
- A mérés az előzőleg bemutatott tptBm benchmark eredményeit mutatja



Index-adatszerkezetek

- Hash-alapú struktúrák
 - nincs markáns különbség
- B*-fa (DRDB)
 - lassú. ?!



(a) egy B-fa csomópont

(b) egy B-fa

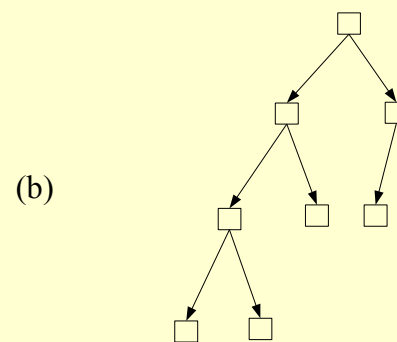
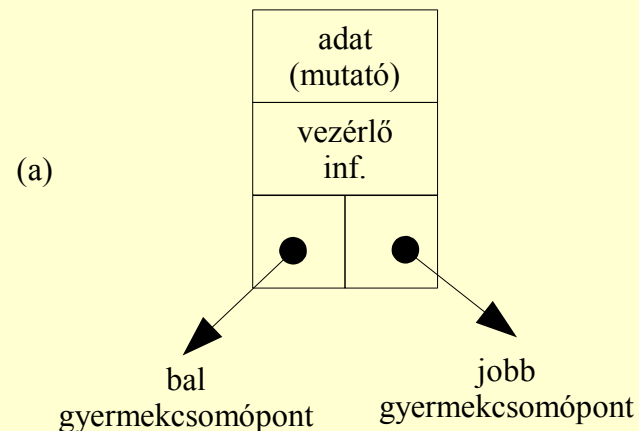


Index-adatszerkezetek (2)

- AVL-fa
 - bináris keresőfa AVL-tulajdonsággal
 - gyenge memóriakihasználás
 - gyors navigáció

(a) egy AVL-fa csomópont

(b) egy AVL fa

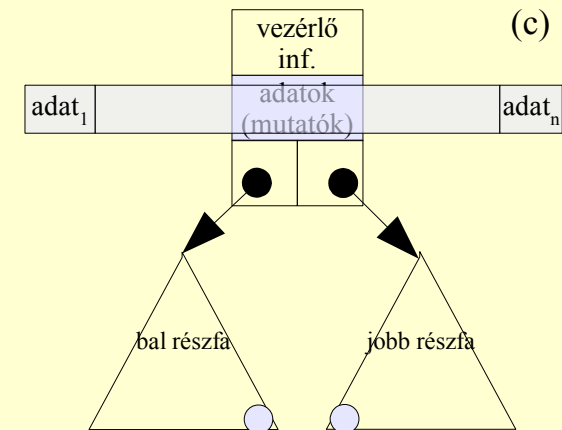
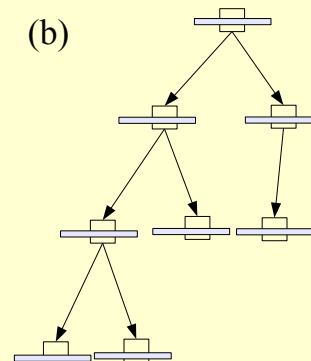
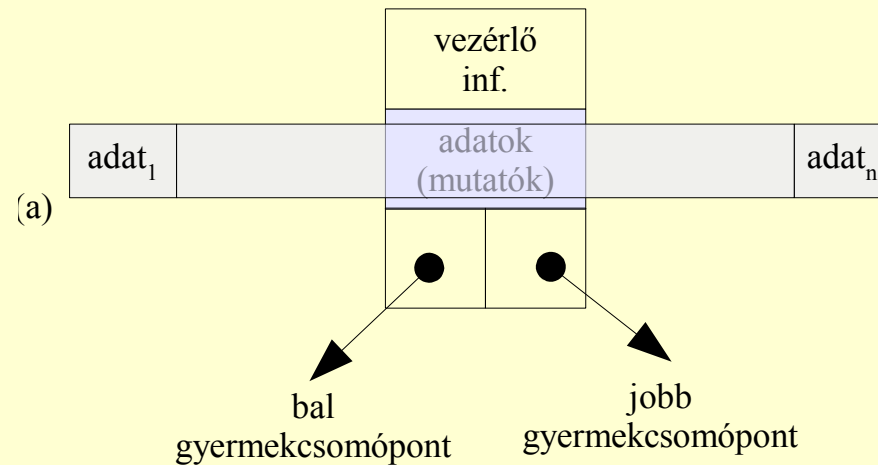




Index-adatszerkezetek: T-fa

(a) egy csomópont;
(b) egy T-fa
(c) a tárolt elemek
elhelyezkedése a T-fában

- B*-fa és AVL-fa előnyei összegyúrva
 - gyors navigáció
 - jó tárkihasználás
- adat tárolása
 - adat maga
 - mutató
 - mutató + adat része (pkT-fa)





Relációk tárolása

- rekordok vagy mutatók tömbje
- idegen kulcsok követésének (illesztés) támogatása
 - avagy „előreszámított (fél)illesztések”
 - DRDB: klaszterezés
 - IMDB: a kulcs értéke helyett mutató



Lekérdezésvégrehajtás

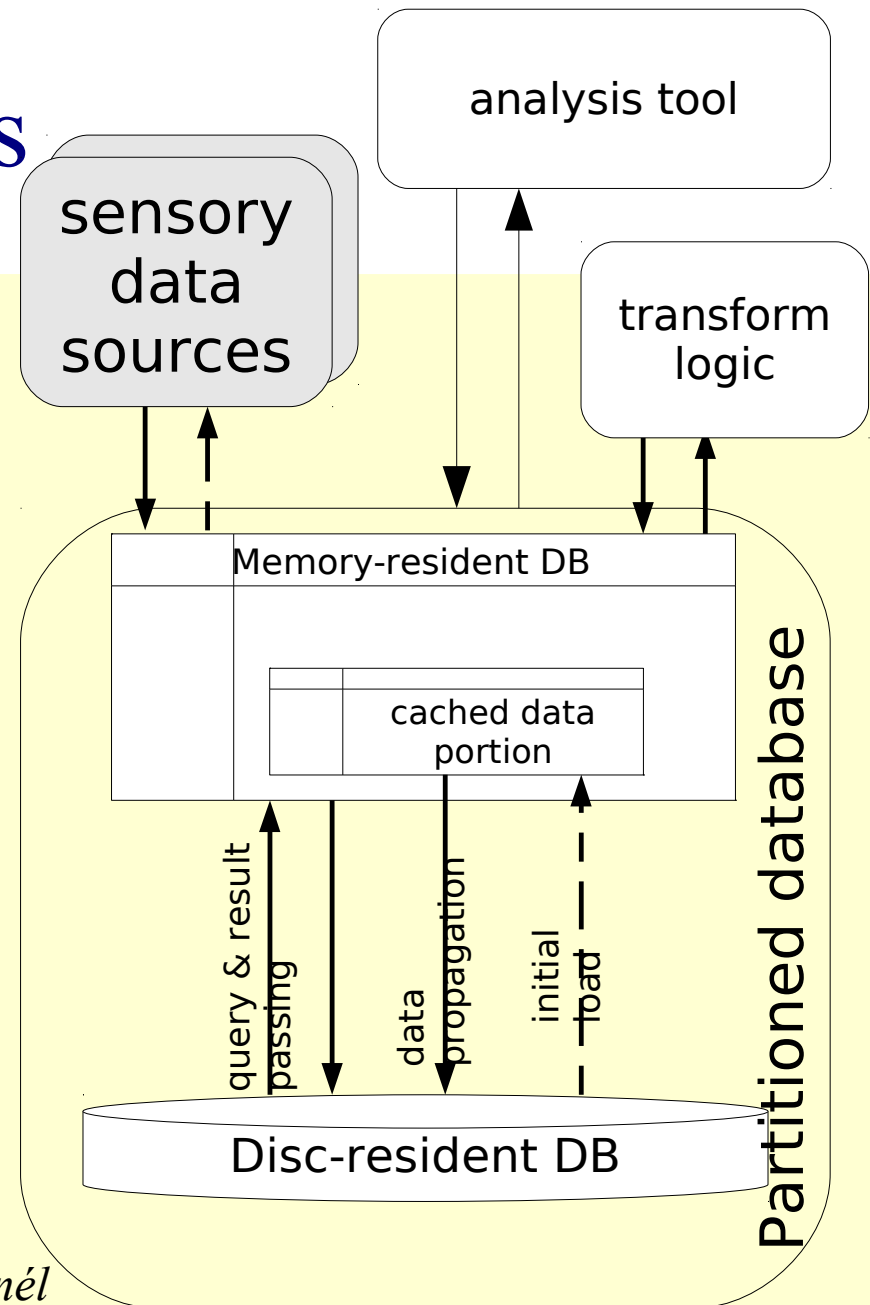
- Vetítés, szelekció:
 - sokszor olcsóbb az eredményreláció leírójának módosítása és on-the-fly számítás, mint a tényleges számítás
- Illesztések:
 - 1. a mutatók követéséről írottakat
- Költség:
 - DRDB: tipikusan az eredmény kiírásának költségét nem számoljuk (miért?), de
 - IMDB: a saját címtér miatt nem mindig merül fel



Particionált adatbázis

- ha a teljes adatbázis IMDB-ben nem
 - megoldható v.
 - racionális
- Hibrid megoldás
 - tiszta IMDB
 - in-memory cache
 - DRDB

Az elemzőeszközök a partícionált adatbázishoz csatlakoznak, amíg az adatforrások friss adatai közvetlenül az IMDB részbe érkeznek, hogy minél hatékonyabb lehessen az előfeldolgozásuk.





Vajon IMDB-nek minősül? (ism)

- egyedi alkalmazás saját adatszerkezetekkel
- hagyományos DBMS úgy, hogy
 - nagy fájlrendszer cache az operációs rendszerben
 - adatfájlok RAM-diszken
 - nagy a buffer cache



Kérdések...

- ...válaszok (?)



Köszönöm megtisztelő figyelmüket!

Marton József Ernő
marton@db.bme.hu



Adatbázisok haladóknak 2012.

2012. szeptember 18.



Demó: durable commit hatása – benchmark

- Az instance `dir/demo/performance` alatt található a TimesTen részeként szállított **tptbm** multi-user throughput benchmark

```
$ ./tptbm -read 20 -insert 40 -xact 100000 -key 1000  
-proc 5 -connstr  
"DSN=TptbmData_tt70;DurableCommits=0;PermSize=400"
```

Connecting to the data source

Populating benchmark data store

Waiting for 5 processes to initialize

Beginning execution with 5 processes: 20% read, 40%
update, 40% insert

Elapsed time: 12.7 seconds

Transaction rate: 39513.2 transactions/second

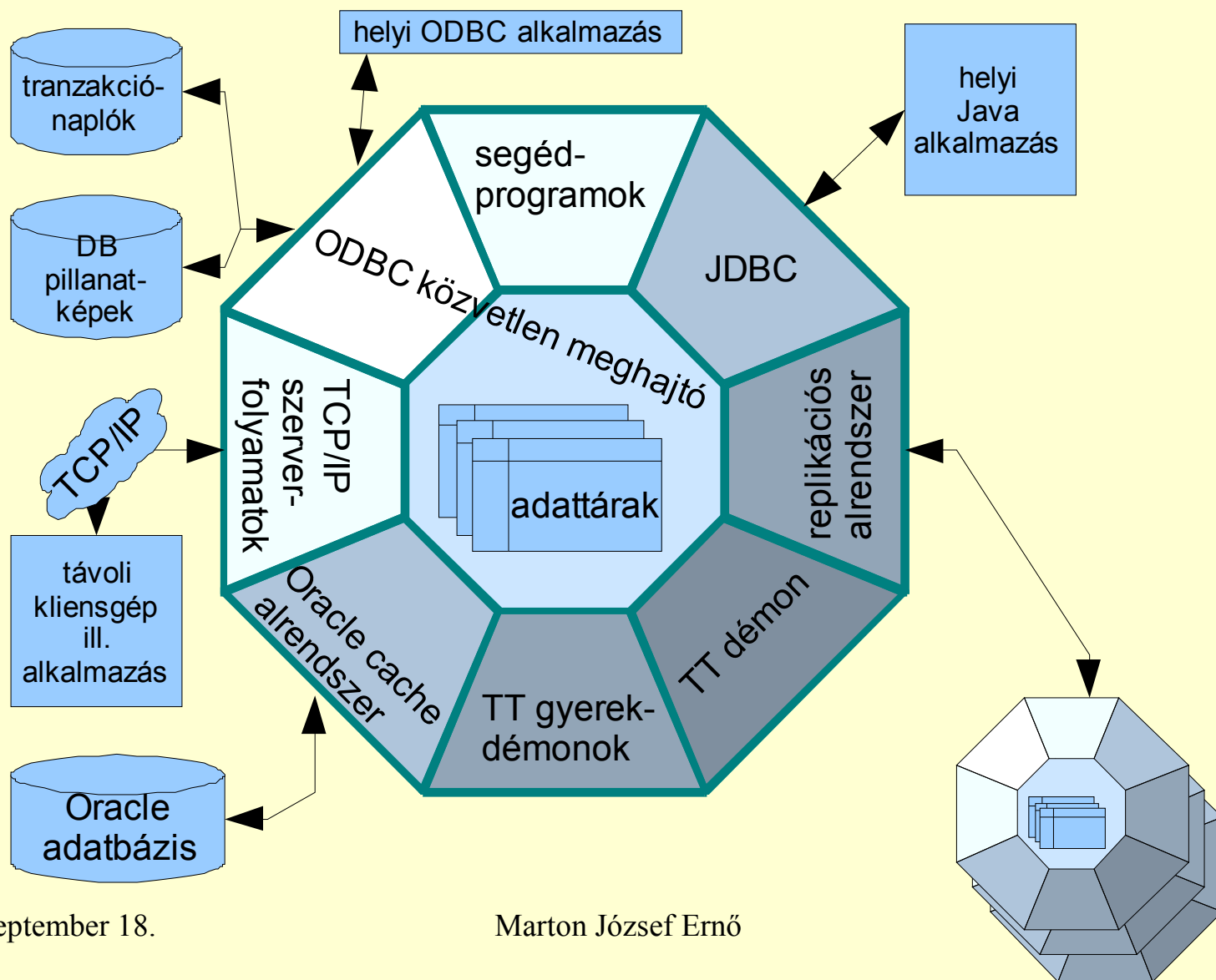


Megvalósítások

- MySQL memory engine
- Oracle TimesTen



Oracle TimesTen rendszerbe illesztése





TimesTen – adatszervezés

- Három szintű hierarchia
 - TimesTen példány (instance)
 - démon, kezelt adatok, segédprogramok (pl. ttIsql)
 - jogosultságkezelési egység
 - kivétel: tulajdonos a sémobjektumaira teljes jogkörű
 - adattár (data store)
 - naplózási és mentési/visszaállítási egység
 - rendelkezésre áll === memóriába töltött
 - séma (schema)
 - felhasználó objektumai (tábla, nézet, szekvencia, ...)



IMDB alkalmazási példa – mobilkommunikáció

- Mobilkommunikáció fejlődése
 - bővülő szolgáltatások, ügyfélkör
 - előrefizetett szolgáltatások: egyenlegfigyelés
- A hálózat cellás, moduláris felépítésű
 - elosztott, valós idejű adatbázis-rendszer az adatgyűjtésre
- Csalásdetekció
 - pl. az ügyfél földrajzi mozgásából (fizikai korlátok), stb.



IMDB alkalmazási példa – tőzsdeinformatika

- Eladási és vételi ajánlatok intelligens illesztése
 - haszon ☺
- Jegyzések adatai, valós idejű elemzések
 - haszon ☺
- NYMEX: 2005 óta memóriaadatbázis-alapú real-time adatfolyamban közlik a jegyzéseket
- A sokszereplős valós idejű információszolgáltatást hozta el nagyságrendileg olcsóbb infrastruktúrára az IMDB