

Négy adatbázis-kezelő rendszer összehasonlítása webes környezetben

Sterbinszky Nóra
snorav@gmail.com

Áttekintés

- ▶ Növekvő igény hatékony adatbázis-kezelőkre a világhálón
- ▶ Hogyan mérhető ezek teljesítménye a weben?
- ▶ Az erre használható TPC webes szabványának bemutatása
- ▶ A tesztelés eredménye

2009. április 24.

INFO Savaria 2009

2

Vizsgált adatbázis-kezelők

- ▶ Apache Derby
- ▶ MySQL
- ▶ Oracle Database
- ▶ PostgreSQL

2009. április 24.

INFO Savaria 2009

3

Az Internet fejlődése

- ▶ Katonai célú hálózat (ARPANET)
- ▶ Elektronikus levél (levelezőlisták, hírcsoportok)
- ▶ Fájltranzfer, távoli bejelentkezés
- ▶ WWW: hang-, kép- és videofájlok
- ▶ Elektronikus kereskedelem
(A háttérben egyre hatékonyabb adatbázisokra van szükség.)

2009. április 24.

INFO Savaria 2009

4

Kétféle szemléletmód: OLTP és OLAP

- ▶ OLTP (On-Line Transaction Processing)
 - A feldolgozandó adatok száz-as nagyságrendűek
 - Hatékony tranzakció-kezelés
 - Olvasási és írási műveletek egyaránt gyakoriak
- ▶ OLAP (On-Line Analytical Processing)
 - Rendkívül hatékony az olvasási műveletek terén
 - A feldolgozandó adatok milliós nagyságrendűek
 - Többdimenziós lekérdezések

2009. április 24.

INFO Savaria 2009

5

Kétféle szemléletmód: OLTP és OLAP

- ▶ A négy vizsgált adatbázis-kezelő mindegyike OLTP típusú az alapbeállításuk szerint
- ▶ Valamilyen módon megoldott az OLAP-funkció használata is (Derby: OLAP-lekérdezések, MySQL és PostgreSQL: külön letölthető modul, Oracle: integrált OLAP opció)
- ▶ OLAP szemléletű adatbázis-kezelők (Monet)

2009. április 24.

INFO Savaria 2009

6

A TPC-W szabvány áttekintése

- ▶ Az alkalmazás
- ▶ Az adatbázis
- ▶ Képek

2009. április 24.

INFO Savaria 2009

7

A TPC-W szabvány

- ▶ A Transaction Processing Performance Council webes szabványa (V1)
- ▶ Olyan alkalmazást definiál, amely az elektronikus kereskedelem forgalmát képes szimulálni
- ▶ Az alkalmazás egy on-line könyvtárház implementációja

2009. április 24.

INFO Savaria 2009

8

A TPC-W szabvány

- ▶ A vásárlók böngésző és vásárló tevékenységének szimulálása biztosítja a valós világnak megfelelő tesztelési környezetet
- ▶ Ennek a környezetnek része a webszerver és az adatbázis-szerver is

2009. április 24.

INFO Savaria 2009

9

A TPC-W szabvány Az alkalmazás

- ▶ Az alkalmazás 14 weboldalból áll
- ▶ A felhasznált disztribúcióban ezek Java Servletekkel vannak megvalósítva
- ▶ A felhasználók tevékenységeinek szimulálása RBE-k (Remote Browsing Emulator) segítségével történik
- ▶ Ezek és a teszteléshez szükséges összes kód Java osztályokban helyezkedik el

2009. április 24.

INFO Savaria 2009

10

A TPC-W szabvány Az alkalmazás

- ▶ A felhasználói viselkedésminták a valóságot (valós forgalmat) tükrözik
- ▶ A szabvány ennek megfelelően három felhasználói profilt definiál:
 - Böngésző (95% olvasás, 5% írás)
 - Vásárló (80% olvasás, 20% írás)
 - Rendelő (50% olvasás, 50% írás)
- ▶ A teljesítmény mérésére a másodpercenkénti webes interakciók száma (WIPS) szolgál

2009. április 24.

INFO Savaria 2009

11

A TPC-W szabvány Az alkalmazás

- ▶ Az alkalmazás főbb jellemzői:
 - Dinamikus oldalgenerálás
 - Adatbázis-elérés a weboldalakról (lekérdezés, módosítás)
 - A párhuzamos működés miatt az adatbázis-kezelő rendszerhez egyszerre futnak be feldolgozásra a változó típusú és bonyolultságú műveletek

2009. április 24.

INFO Savaria 2009

12

A TPC-W szabvány Az adatbázis

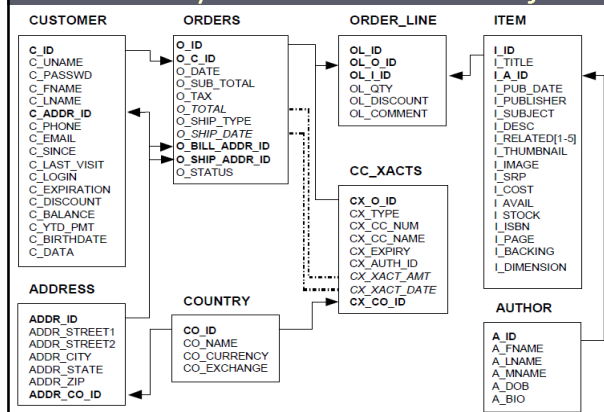
- ▶ Az adatbázis mérete változtatható bizonyos keretek között (kicsi, közepes, nagy)
- ▶ A tesztelés közepes adatbázison történt
- ▶ Az adatbázis sémáját szigorúan rögzíti a szabvány:
 - Táblák
 - Attribútumok
 - Kulcsok
 - Kapcsolatok (relációk)

2009. április 24.

INFO Savaria 2009

13

A szabvány által előírt adatbázis sémája



A TPC-W szabvány Az adatbázis

- ▶ Egymástól függő táblák mérete

Tábla	Sorok száma
AUTHOR	0.25 * ITEM
CUSTOMER	2880 * RBE-k száma
ADDRESS	2 * CUSTOMER
ORDERS	0.9 * CUSTOMER
ORDER_LINE	3 * ORDERS
CC_XACTS	1 * ORDERS

2009. április 24.

INFO Savaria 2009

15

A TPC-W szabvány Az adatbázis

- ▶ Bizonyos attribútumok véletlenszerűen generált karaktersorozatokról és számokról állnak
- ▶ Más attribútumok létező dolgokat képviselnek (pl. az országok létező országok)
- ▶ A valós világban létező dolgok közül véletlenszerűen történik a konkrét attribútum-érték kiválasztása

2009. április 24.

INFO Savaria 2009

16

A TPC-W szabvány Képek

- ▶ A képek méretét szintén meghatározza a szabvány: 5K, 10K, 50K, 100K és 250K nagyságúak lehetnek.
- ▶ Ajánlott, hogy a különböző méretű képek mindegyike szerepeljen az alkalmazásban
- ▶ A felhasznált disztribúcióban a legnagyobb és legkisebb méretűek szerepelnek

2009. április 24.

INFO Savaria 2009

17

A TPC-W szabvány Képek

- ▶ Minden képhez kell, hogy tartozzon thumbnail, amelyek mérete egységesen 5K
- ▶ A generált képek darabszámát mi határozhatjuk meg, de ajánlott annyi darabnak lennie, ahány ITEM van
- ▶ Ebben az esetben ez 10000 képet jelent, valamint ugyanennyi thumbnail-t
- ▶ 13 kép fix (pl. a TPC-W logója)

2009. április 24.

INFO Savaria 2009

18

Mérési környezet

- ▶ A számítógép specifikációja:
 - Hardver: Dell Inspiron 1525 laptop
 - Processzor: Intel^(R) Core^(TM)2 Duo CPU T8100 2.10 GHz
 - RAM: 2 GB
 - Operációs rendszer: Microsoft Windows VistaTM Home Basic Service Pack 1

2009. április 24.

INFO Savaria 2009

19

Mérési környezet

- ▶ Adatbázis-szerverek:
 - Apache Derby 10.4.2.0
 - MySQL 6.0.9
 - Oracle Database 11g
 - PostgreSQL 8.3.3
- ▶ Webszerver:
 - Apache Tomcat 6.0.18

2009. április 24.

INFO Savaria 2009

20

Mérési környezet

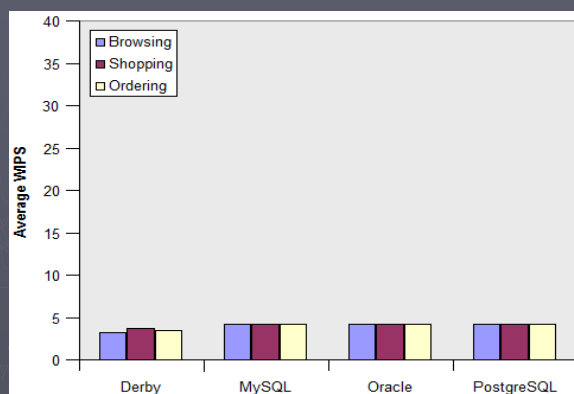
- ▶ A teszteléshez szükséges különböző mértékű terhelést az EB-k (Emulated Browsers) száma és a három, felhasználói viselkedésmintákat tartalmazó profil adta
- ▶ EB-k száma: 30, 50, 100, 150, 300 és 600
- ▶ Felhasználói profilkok: böngésző, vásárló, rendelő

2009. április 24.

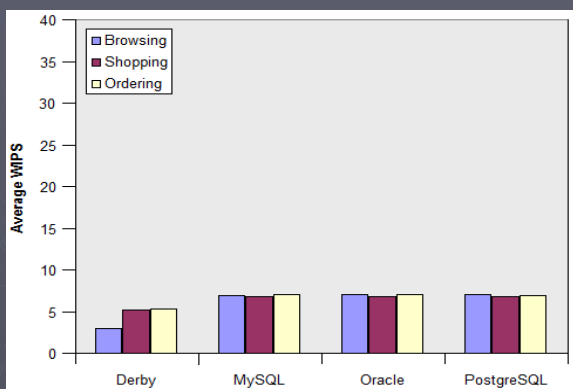
INFO Savaria 2009

21

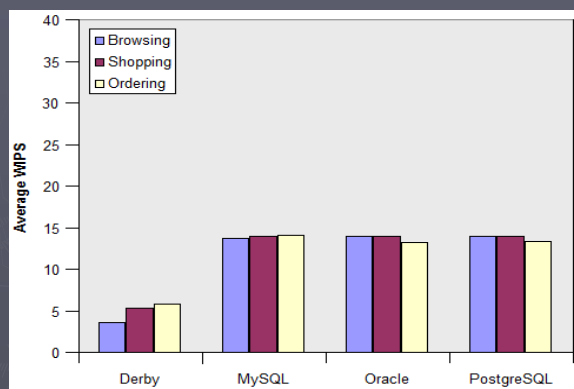
Tesztelés 30 RBE-vel



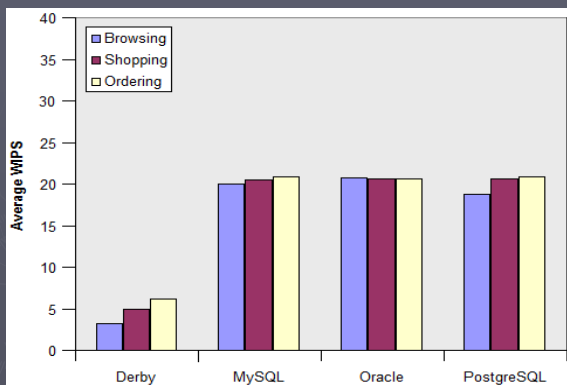
Tesztelés 50 RBE-vel



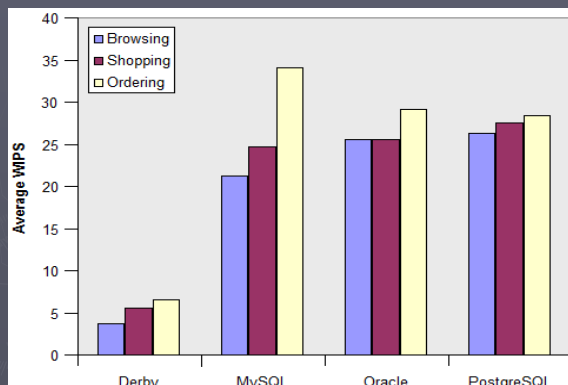
Tesztelés 100 RBE-vel



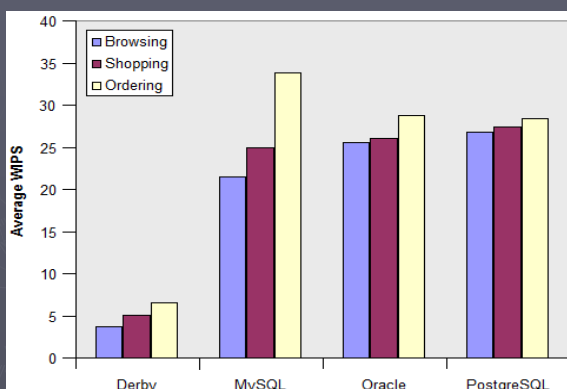
Tesztelés 150 RBE-vel



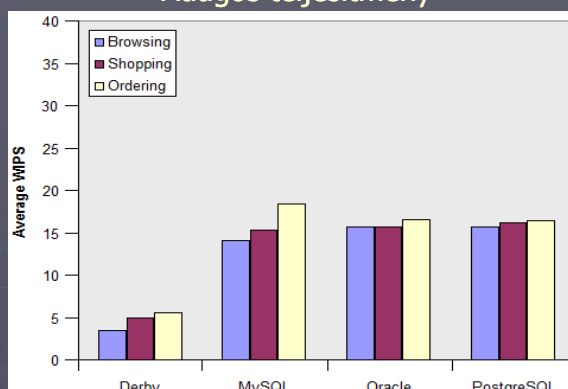
Tesztelés 300 RBE-vel



Tesztelés 600 RBE-vel



Átlagos teljesítmény



Összegzés

- ▶ A Derby alacsony teljesítményt nyújtott, és 100 RBE fölött nem volt fokozható a hatékonysága
- ▶ A másik három adatbázis-kezelő nagyjából a várt eredményt nyújtotta
- ▶ A MySQL különösen jónak tűnik írási műveletek terén, jobbnak, mint az Oracle vagy a PostgreSQL

Összegzés

- ▶ Utóbbi kettő viszont megelőzi a MySQL-t, ha olvasási műveletekről van szó
- ▶ Az RBE-k számának emelésével ezek az eltérések egyre szignifikánsabbak
- ▶ Nagy forgalom mellett a MySQL, Oracle vagy PostgreSQL egyike lehet jó választás

Források

- <http://www.tpc.org/tpcw/default.asp>
- Transaction Processing Performance Council (TPC) Benchmark™ W (Web Commerce) Specification Version 1.8, Feb 19, 2002
- <http://www.ece.wisc.edu/~pharm/tpcw.shtml>
- http://info.ilab.sztaki.hu/~lukacs/AdatbanyaEA2005/dm05_04_adattarhaz.pdf
- <http://en.wikipedia.org/wiki/OLTP>
- http://en.wikipedia.org/wiki/Online_analytical_processing

2009. április 24.

INFO Savaria 2009

31

Köszönöm a figyelmet!

Elérhetőség: Sterbinszky Nóra
snorav@gmail.com

2009. április 24.

INFO Savaria 2009

32