



Adattárház specifikus RDBMS funkciók MPP implementáció

DWH @ Adatb haladóknak

Lévai Ákos
alevai@prise.hu



Adatbázisok haladóknak 2012.

2012. október 30.



Miről lesz szó? (1p)

Adattárház és céljai

Üzleti és műszaki elvárások

MPP RDBMS

Egy implementáció: Teradata

Ajánlott irodalom / Keresések



Adattárház és céljai (5p)

- Mi az adattárház? (Inmon/Kimball)
- Célok
 - A döntéshozókat **időben** ellátni információkkal
 - R&D elemzések támogatása
 - Aktív részvétel a proaktív üzletmenetben
 - „Single version of truth”



Néhány elvárás – Megoldások (5p)

- Előre definiált és ad-hoc kérdések megválaszolása
 - Hosszú időszak lefedése
 - Feladatnak megfelelő válaszüidő
 - Kevert „workload” kezelése
 - Hosszabb táv: Skálázhatóság
-
- **Célra optimalizált Applikáció**
 - **Célra optimalizált (R)DBMS**



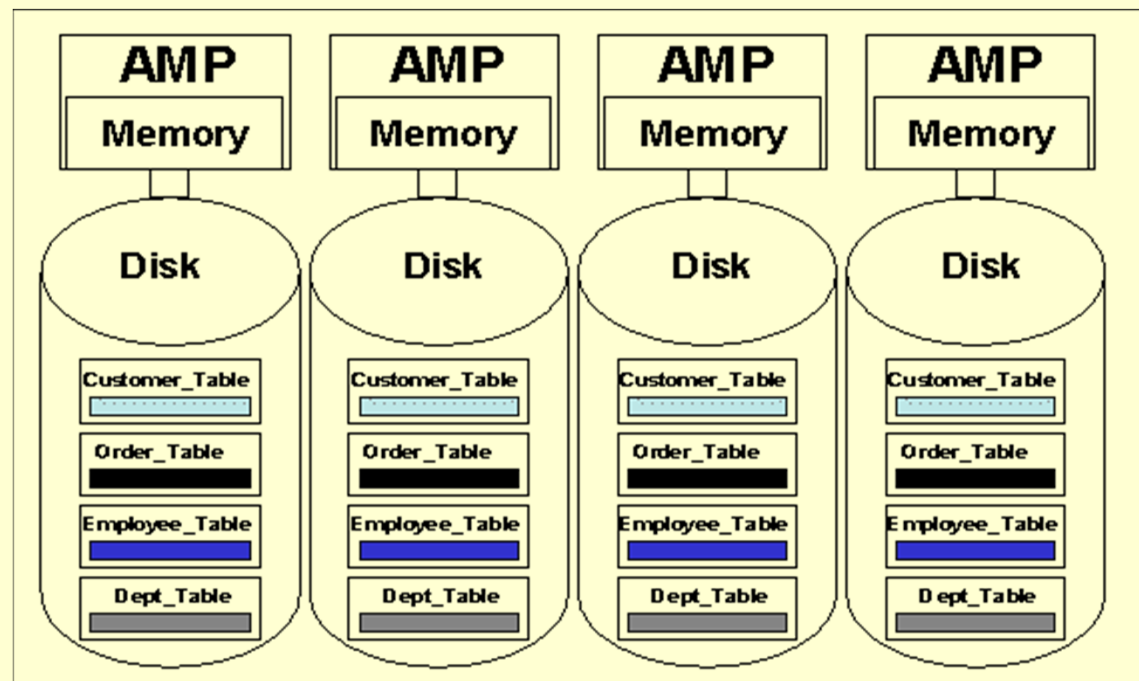
Néhány technika (3p)

- DB-n kívüli párhuzamos processzálás (Hadoop, MapReduce)
- Denormalizált Logikai-Fizikai adatmodellezés
- „Multi temperature” adattárolás
- Dinamikus resource management
- **Párhuzamosítás – MPP architektúra**



MPP RDBMS architektúra (6p)

- Egyforma unit-ok
- Feladat (lehetőleg) egyenlő felosztása
- Implementációk: Greenplum, Netezza, Teradata
- Tipikusan share nothing felépítés (pl Teradata [1])





Adattárház feladatelemek (5p)

A tipikus RDBMS mumusok

- Full table scan
 - Összegzések,összetett keresések
- Sort
 - Analitikai műveletek,indexfelépítés
- Join

Nem kell tőlük félni,hatékonyan párhuzamosíthatók!



Teradata aspektus (10p)

- Felhasználó felé transzparens párhuzamosítás
 - Hashing
- Tervezési irányelvek
 - „**Know data**” – „Know workload”
 - Fő feladat: megfelelő disztribúció
 - Nem mindig a legegyszerűsebb az optimális!
 - Indexelés: ritka
 - Cost based optimizer: tábla-mező statisztikák létfontosságúak



MPP RDBMS előnyök (5p)

- „Olcsó” SMP építőelemek
 - SQL interface – általánosan használható
 - Optimális esetben $\sim 1/N^*$ futásidő
 - Lineáris skálázhatóság! (>1000 unit bizonyítva)
 - Dinamikus bővíthetőség
-
- *N: unitok száma



Kérdések...

DVH @ Adatb haladóknak



2012. október 30.

Lévai Ákos

10.



GoogleWords - Ajánlott irodalom

[MPP RDBMS, Teradata, Greenplum, Big Data, Hadoop, MapReduce, Teradata Express](#)

http://scs.web.elte.hu/Work/DW/papers/sidlo_dipl.pdf

<http://www.cubrid.org/blog/dev-platform/platforms-for-big-data/>

<http://www.teradata.com/teradata-express/>

Források

[1]

http://www.coffingdw.com/Teradata_Basics/chapter_9___advanced_topics_that_you_will_be_tested_on/teradata_is_a_shared_nothing_architecture.htm



Köszönöm a figyelmet!

DVH @ Adatb haladóknak

Lévai Ákos
alevai@prise.hu



Adatbázisok haladóknak 2012.

2012. október 30.