



Oszlopalapú adatbázis-kezelők

Barabás Ákos
akos.barabas@gmail.com



Adatbázisok haladóknak 2012. ősz

2012. október 9.

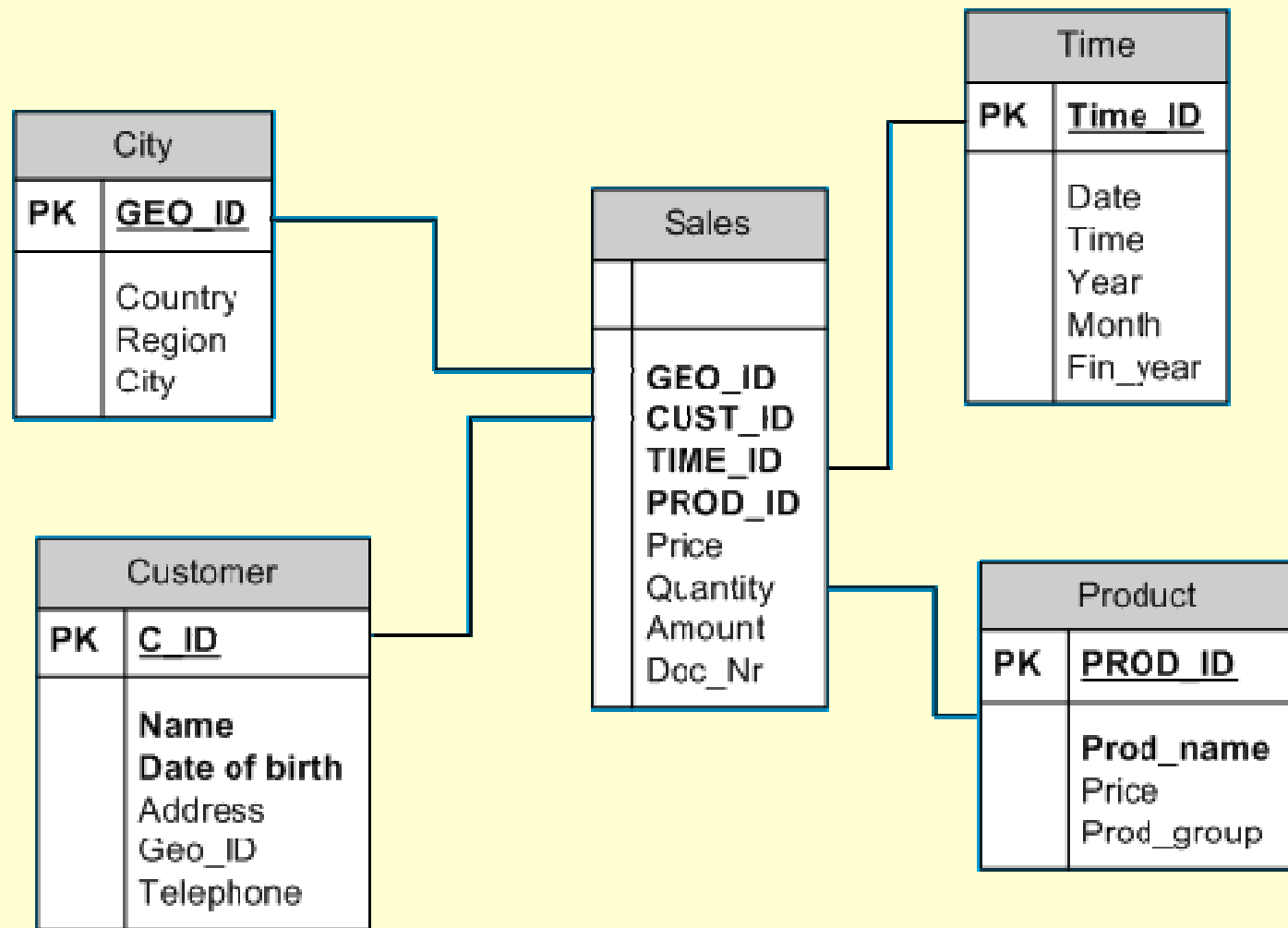


Agenda

1. Motiváció
2. Oszlopalapú szervezés, lekérdezés kiértékelés
3. Előnyök, hátrányok
4. Teljesítmény
5. Tömörítési módszerek
6. Kiegészítő technikák
7. Piaci kitekintés

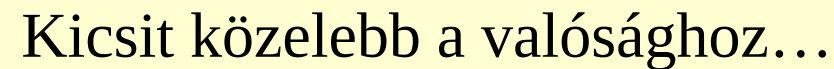


Motiváció





CODB @ Adate haladóknak





Motiváció

```
SELECT d.name, COUNT(f.time_id),  
      AVG(f.amount)  
FROM sales AS f, customer AS d,  
      time AS t  
WHERE f.cust_id = d.c_id  
AND f.time_id = t.time_id  
AND d.address = 'Futrinka utca'  
AND t.year = 2011  
GROUP BY d.name;
```



Oszlopalapú szervezés – Alapok I.

ID	Név	Életkor	Lakhely
1000	Klemens	42	Stuttgart
1001	Rajesh	29	Delhi
1002	Francesco	30	Rome
1003	Colin	51	Dublin

Soralapú (klasszikus) szervezés:

1. blokk	1000	Klemens	42	Stuttgart
2. blokk	1001	Rajesh	29	Delhi
3. blokk	1002	Francesco	30	Rome
4. blokk	1003	Colin	51	Dublin



Oszlopalapú szervezés – Alapok II.

ID	Név	Életkor	Lakhely
1000	Klemens	42	Stuttgart
1001	Rajesh	29	Delhi
1002	Francesco	30	Rome
1003	Colin	51	Dublin

Oszlopalapú szervezés:

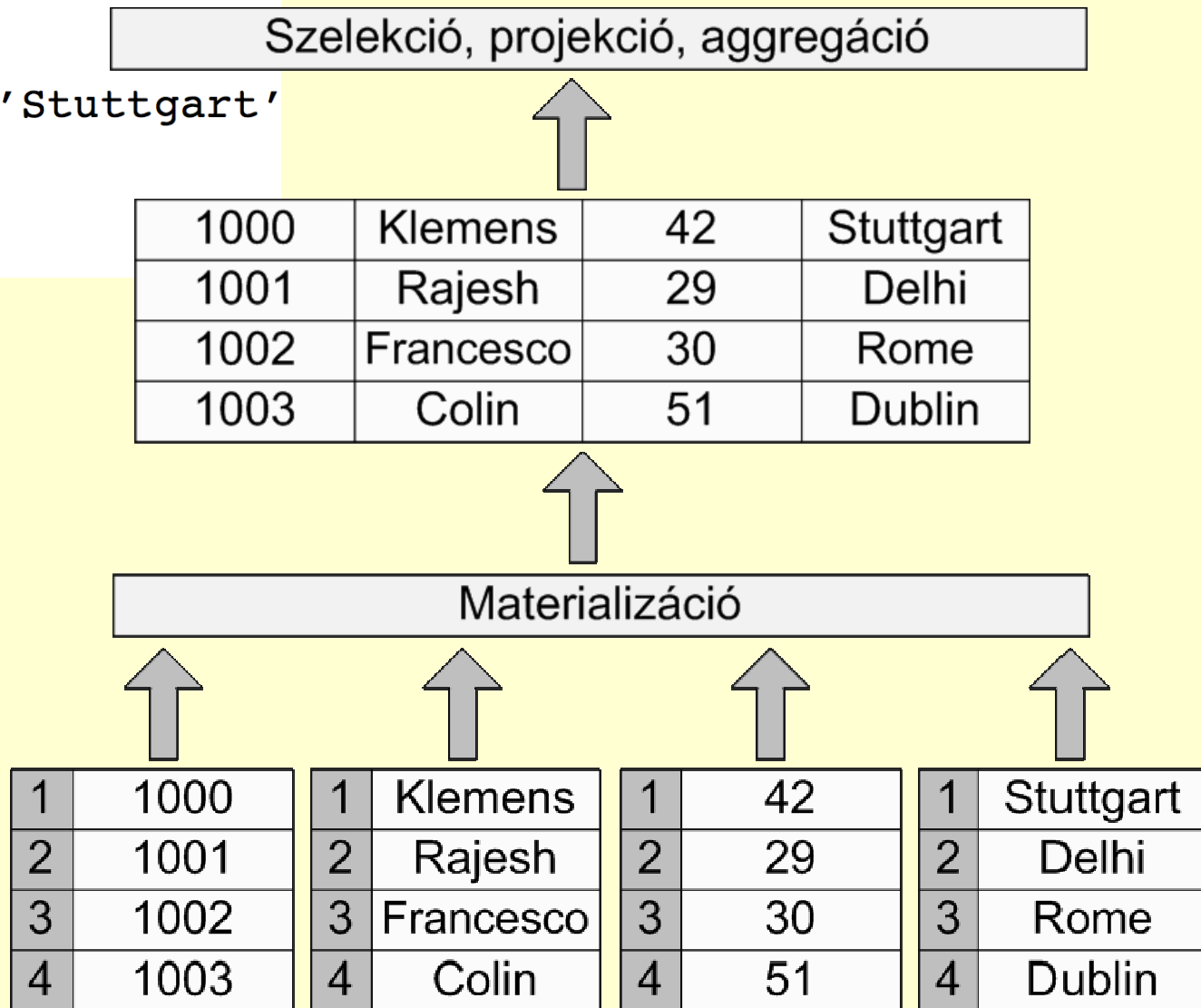
1. blokk	1	1000	2	1001	3	1002	4	1003
2. blokk	1	Klemens	2	Rajesh	3	Francesco	4	Colin
3. blokk	1	42	2	29	3	30	4	51
4. blokk	1	Stuttgart	2	Delhi	3	Rome	4	Dublin

Továbbra is relációs adatmodell,
csak a fizikai szervezés más!



Oszlopalapú szervezés – Kiértékelés I.

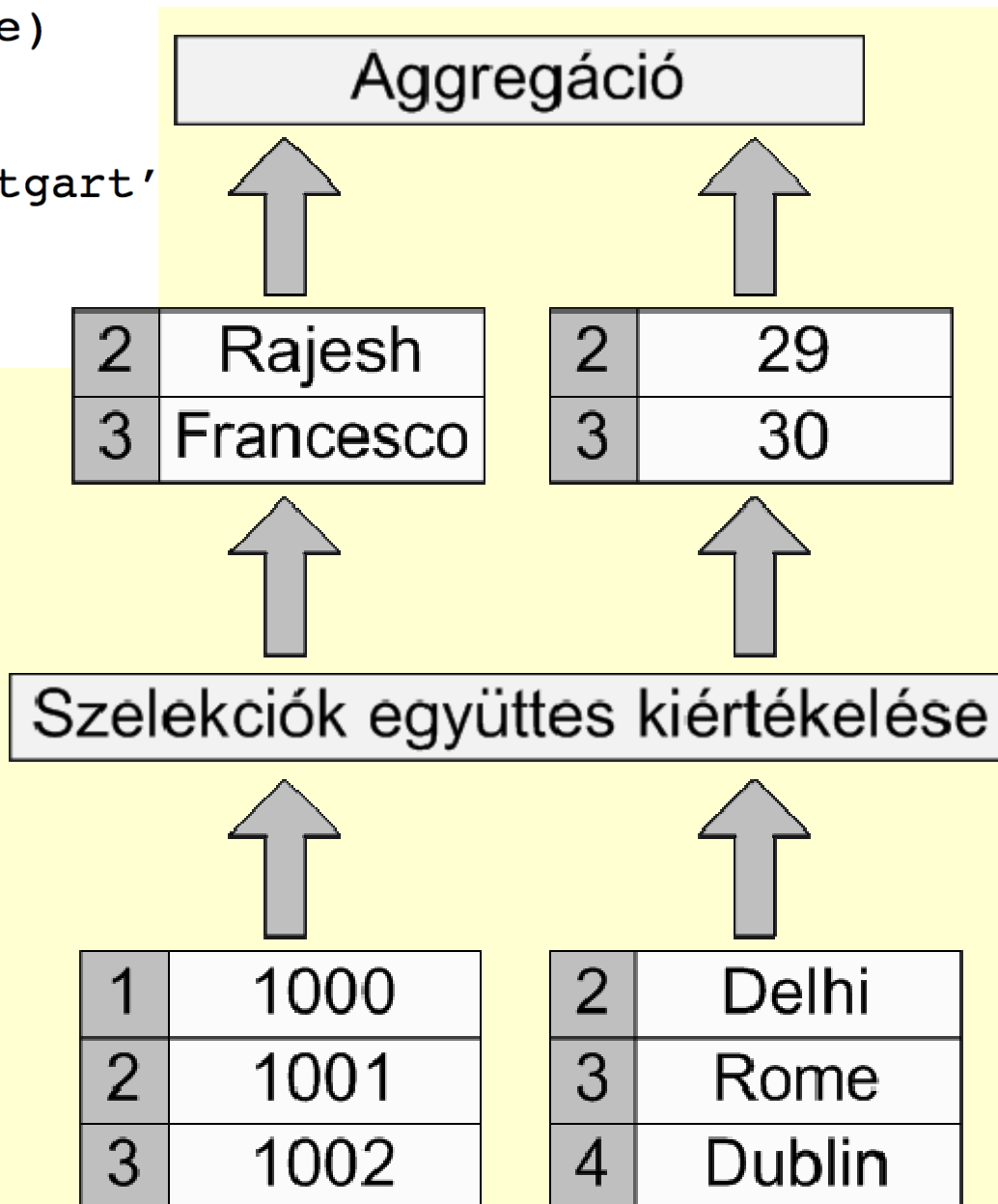
```
SELECT name, MAX(age)
FROM person
WHERE city <> 'Stuttgart'
AND id < 1003
GROUP BY name;
```





Oszlopalapú szervezés – Kiértékelés II.

```
SELECT name, MAX(age)
FROM person
WHERE city <> 'Stuttgart'
AND id < 1003
GROUP BY name;
```





Előnyök/hátrányok

- Előnyök:
 - Nagy adatmennyiség esetén
 - Nagy projektivitás (szelektivitás) esetén
 - Aggregáció
 - Párhuzamosság
 - Hatékony erőforráskihasználás
- Hátrányok:
 - Írási műveletek lassúsága



Teljesítmény

```
SELECT SUM  
(lo_extendedprice*lo_discount) AS revenue  
FROM lineorder, dwdate  
WHERE lo_orderdate = d_datekey  
AND d_year = 1993  
AND lo_discount between  
1 and 3  
AND lo_quantity < 25;
```

Q1

- TPC-H
- Standard csillagséma, lekérdezések
- TPC toplista kissé mást mutat...

Metrika/DB	Soralapú RDBMS	Oszlopalapú RDBMS
Q1 (sec)	32.0	3.4
Q5 (sec)	26.1	3.2
Q8 (sec)	4.1	3.2
Q1-Q12 átlag (sec)	13.6	1.8
Tárterület (GB)	60.3	36.3



Tömörítési módszerek – RLE

- Futamhossz kódolás (run-length encoding)

ID	Keresztnév	Életkor	Lakhely	Nem	Foglalkozás
1000	Klemens	42	Stuttgart	Férfi	informatikus
1001	Rajesh	29	Delhi	Férfi	informatikus
1002	Francesco	30	Rome	Férfi	labdarúgó
1003	Colin	51	Dublin	Férfi	kocsmáros
1004	Mária	46	Budapest	Nő	informatikus

Nem
(Férfi, 1,4)
(Nő, 5, 5)



Tömörítési módszerek – Szótár alapú kódolás

ID	Keresztnév	Életkor	Lakhely	Nem	Foglalkozás
1000	Klemens	42	Stuttgart	Férfi	informatikus
1001	Rajesh	29	Delhi	Férfi	informatikus
1002	Francesco	30	Rome	Férfi	labdarúgó
1003	Colin	51	Dublin	Férfi	kocsmáros
1004	Mária	46	Budapest	Nő	informatikus

Kód	Leírás
1	informatikus
2	labdarúgó
3	kocsmáros

Foglalkozás
1
1
2
3
1



Tömörítési módszerek – Delta tömörítés

ID	Keresztnév	Életkor	Lakhely	Nem	Foglalkozás
1000	Klemens	42	Stuttgart	Férfi	informatikus
1001	Rajesh	29	Delhi	Férfi	informatikus
1002	Francesco	30	Rome	Férfi	labdarúgó
1003	Colin	51	Dublin	Férfi	kocsmáros
1004	Mária	46	Budapest	Nő	informatikus

Életkor
42
29
1
21
46

- Egy változat: frame of reference encoding



Tömörítési módszerek – LZW tömörítés

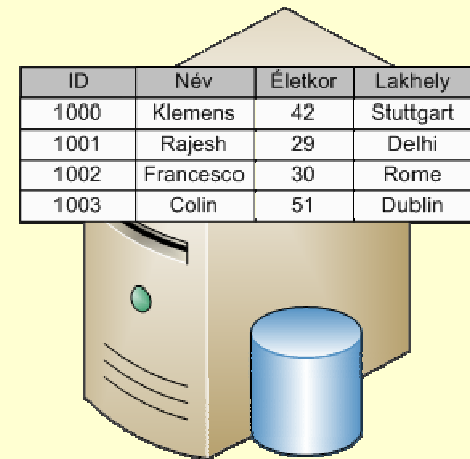
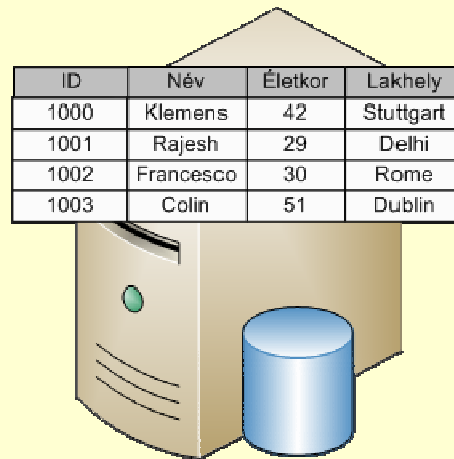
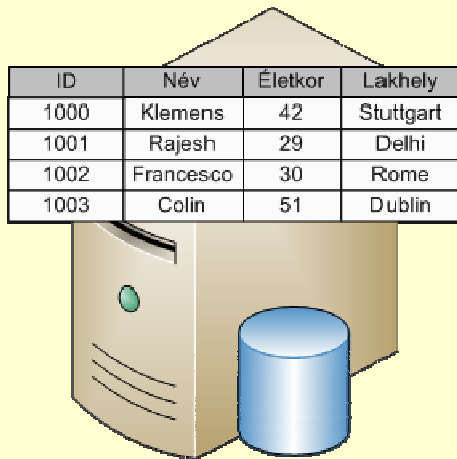
- Lempel-Ziv-Welch (LZW tömörítés)

Lépés	Leghosszabb s a szótárban	Következő karakter	Output	Új szótárbejegyzés (kód – érték)	Szótár tartalma
1	T	R	20	100 – TR	1-26 (A-Z)
2	R	A	18	101 – RA	1-26, 100
3	A	L	1	102 – AL	1-26, 100-101
4	L	A	12	103 – LA	1-26, 100-102
5	AL	A	102	104 – ALA	1-26, 100-103
6	A	–	1	-	1-26, 100-104

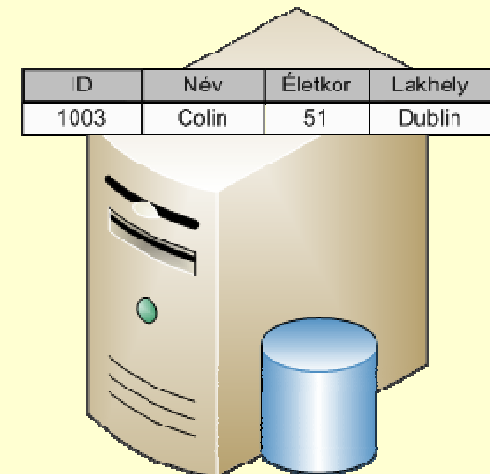
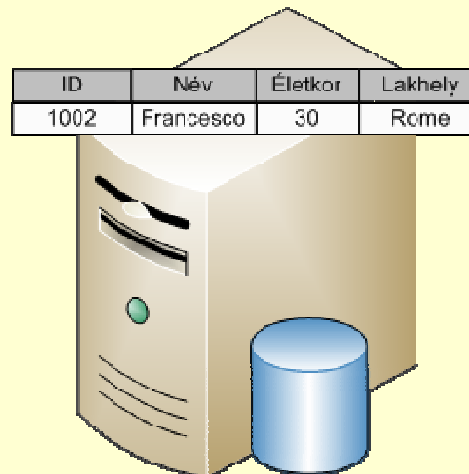
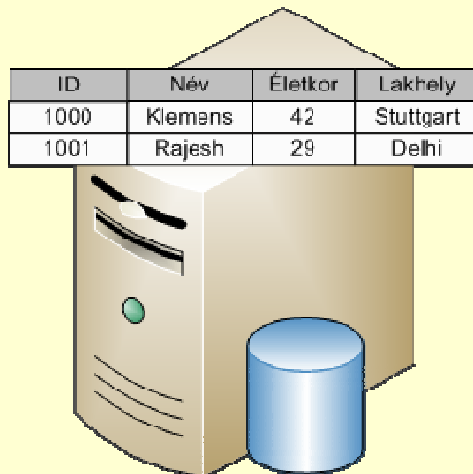


Kiegészítő technikák – Elosztott tárolás

- Replikáció



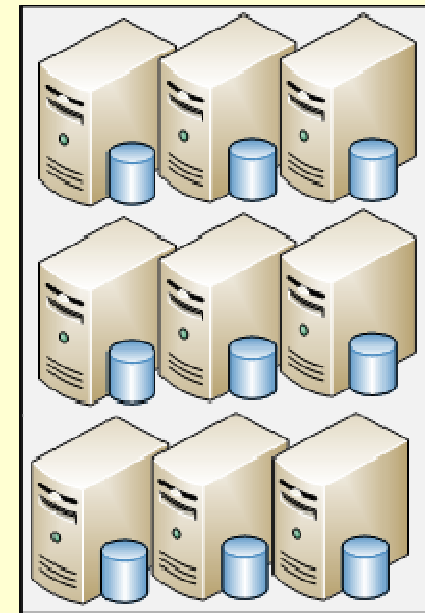
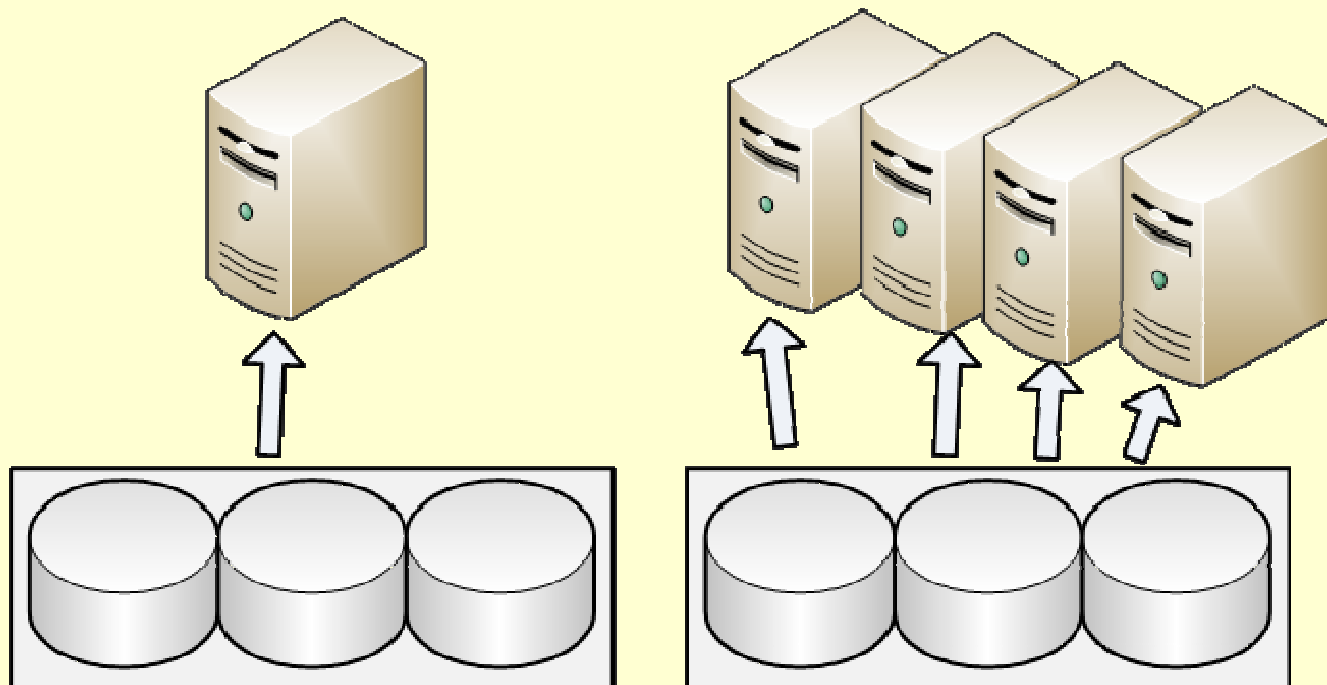
- Sharding





Kiegészítő technikák – MPP

- Hagyományos: belső I/O ☹
- Megosztott diszk: hatékonyság ☹
- Shared nothing: kompakt komponensek, gyors köztes I/O





Kiegészítő technikák – Bitmap index

Keresztnév	Életkor	Nem		Foglalkozás		
		Férfi	Nő	informatikus	labdarúgó	kocsmáros
Klemens	42	1	0	1	0	0
Rajesh	29	1	0	1	0	0
Francesco	30	1	0	0	1	0
Colin	51	1	0	0	0	1
Mária	46	0	1	1	0	0

Projekció/ Szelekció	Klemens	Rajesh	Francesco	Colin	Mária
Nő	0	0	0	0	1
Informatikus	1	1	0	0	1
ÉS	0	0	0	0	1



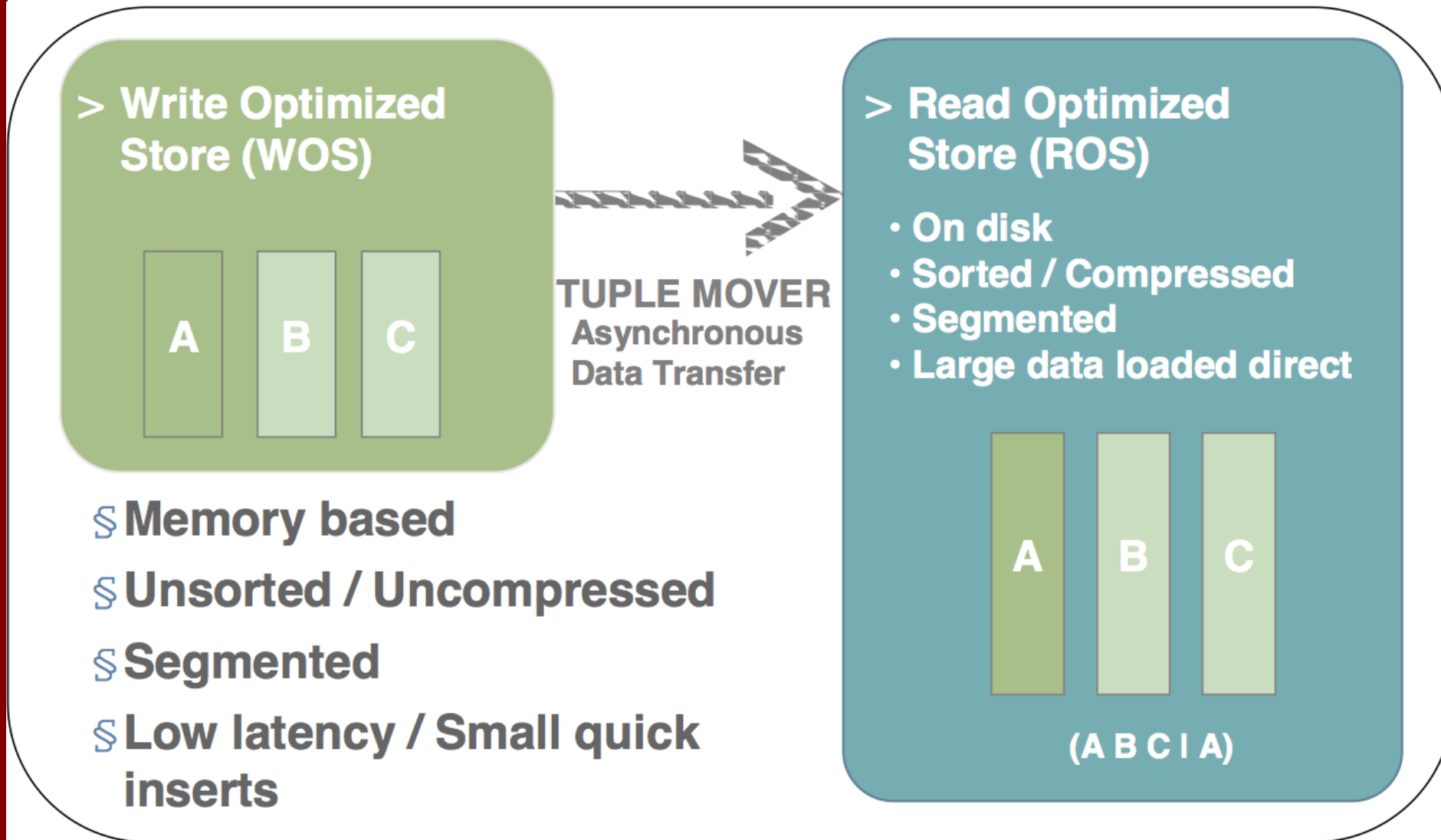
Kiegészítő technikák – Sűrű csomagolás

CODB @ Adatb haladóknak





Kiegészítő technikák – Parallel L&Q





Piaci kitekintés

- „karakterek” a magic quadrant alapján



As of February 2012



Kérdések

CODB @ Adatb haladóknak

?