

МОДЕЛИРАНЕ И АНАЛИЗ НА СОФТУЕР

Павел Кюркчиев

Ас. към ПУ „Паисий Хилендарски“

<https://github.com/pkyurkchiev>

@pkyurkchiev

МНОГОИЗМЕРНА БАЗА ДАННИ

Склад за данни

Многоизмерна база данни

Multidimensional database (MDB)

- Многоизмерната база данни е специфичен тип база данни, която е оптимизирана за складиране на данни и онлайн аналитична обработка (OLAP). Многоизмерната база данни е структурирана от комбинацията на данни от различни източници, които работят едновременно между базите данни и предлагат мрежи, йерархии, масиви и други методи за форматиране на данни. В многоизмерна база данни данните се представят на потребителите чрез многомерни масиви и всяка отделна стойност на данните се съдържа в клетка, която може да бъде достъпна чрез множество индекси.

- Многоизмерната база данни използват концепцията за куб данни (наричана още хиперкуб), за да представят измеренията на данните, които са достъпни за потребителите. Концепцията на многоизмерната база данни е предназначена да подпомага системите за вземане на решения. Тази подробна организация на данните позволява разширено и сложно генериране на заявки, като същевременно осигурява изключителни резултати в определени случаи в сравнение с традиционните релационни структури и бази данни. Този тип база данни обикновено е структурирана в ред, който оптимизира OLAP и приложения за съхранение на данни.

Типове Многоизмерни бази данни

- Електронни таблици (Spreadsheets)
- Основа таблица (Pivot table)
- Кубове (Cubes)
- ...

Факти, Измерения (Величины),
Мерки и Заявки

Факти (Facts)

- Фактите представляват темата - интересният модел или събитие в предприятието, което трябва да бъде анализирано, за да се разбере поведението му. В повечето многоизмерни модели от данни фактите се дефинират имплицитно от тяхната комбинация от стойности на величините; фактът съществува само ако има непразна клетка за определена комбинация от стойности.

- Факт е пример за конкретно събитие или събитие и свойствата на събитието, които са съхранени в база данни.

Продадохте ли си часовника на клиент миналия петък? Това е факт.

Получихте ли доставка на 76 пакета вчера? Това е друг факт.

Типове факти

- Събития (Events)
- Образци (Snapshots)
- Кумулативни моменти образци (Cumulative Snapshots)

Събития (Events)

- Събития, най-малката единица, обикновено моделират събития от реалния свят, като един факт представлява един и същ случай на основно явление.

Примери: Включват продажби, кликвания върху веб страница или движение на стоки във и извън склад.

Образци (Snapshots)

- Образците моделират състоянието на обект в даден момент от време, като например нивата на инвентаризация на склад или броя на потребителите на уеб страница.

Пример от реалния свят: Консервна кутия със зърна на рафта - може да се появи в няколко факта в различни точки от време.

Кумулативни моментни образци (Cumulative Snapshots)

- Кумулативните моментни образци обработват информация за активността до определен момент.

Например: Общите продажби до определен период на тази година могат лесно да бъдат сравнени с цифрите за съответния период на миналата година.

Измерения (Dimensions, Величина)

- Измеренията са съществена и отличителна част от многоизмерните бази данни. Важна цел на многоизмерното моделиране е да се използват измерения, които да осигурят възможно най-голямо описание(значение) на фактите.

- Едно измерение е ключово описание (индекс), чрез който можем да получим достъп до факти според желаната стойност (или стойности).

Можем да организираме данните за продажбите си според тези измерения: време, клиент и продукт.

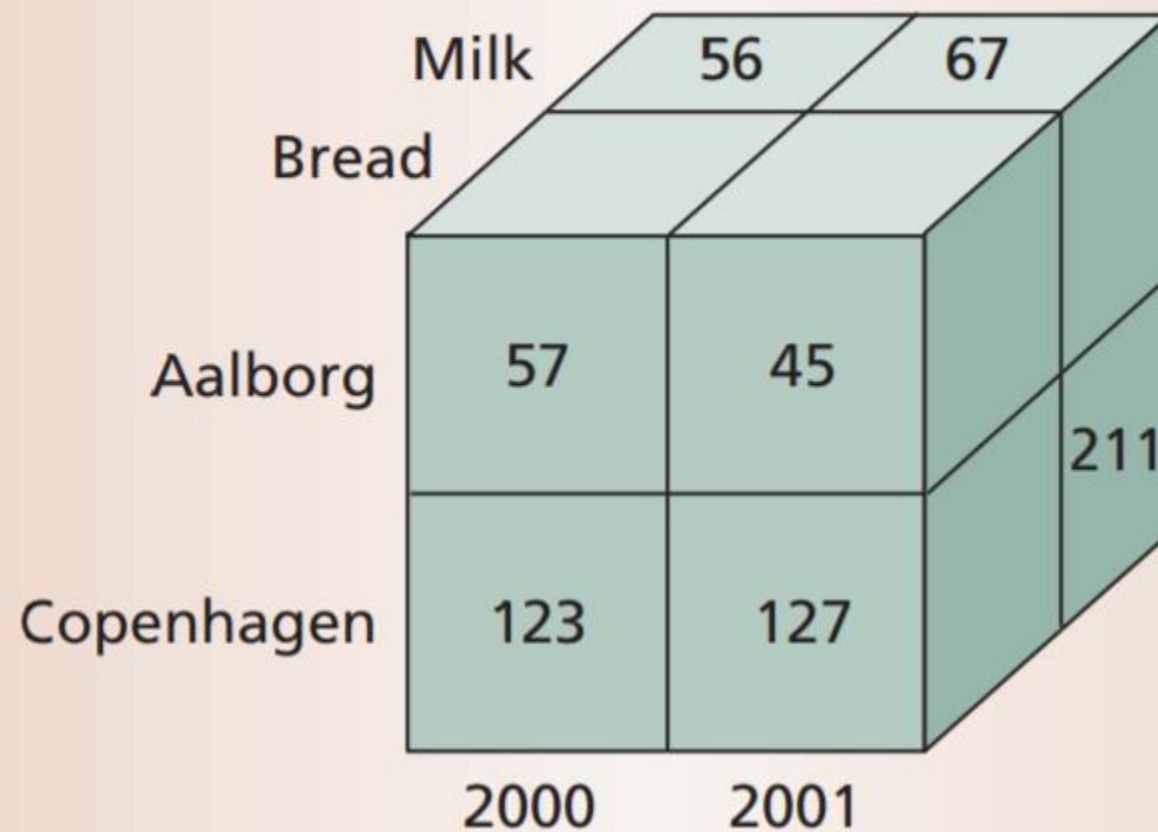
Мерки (Measures)

- В многоизмерна база данни мерките обикновено представляват свойствата на факта, които потребителят иска да оптимизира. Мерките представляват различни стойности за различните комбинации от измерения. Стойността и формулата са избрани да осигурят значима стойност за всички комбинации от нивата на агрегация. Тъй като метаданните определят формулата, данните не се възпроизвеждат както в електронна таблица.

Данни за продукти към фигура 1

Product	Number of purchases by city			
	Aalborg	Copenhagen	Los Angeles	New York City
Milk	123	555	145	5,001
Bread	102	250	54	2,010
Jeans	20	89	32	345
Light bulbs	22	213	32	9,450

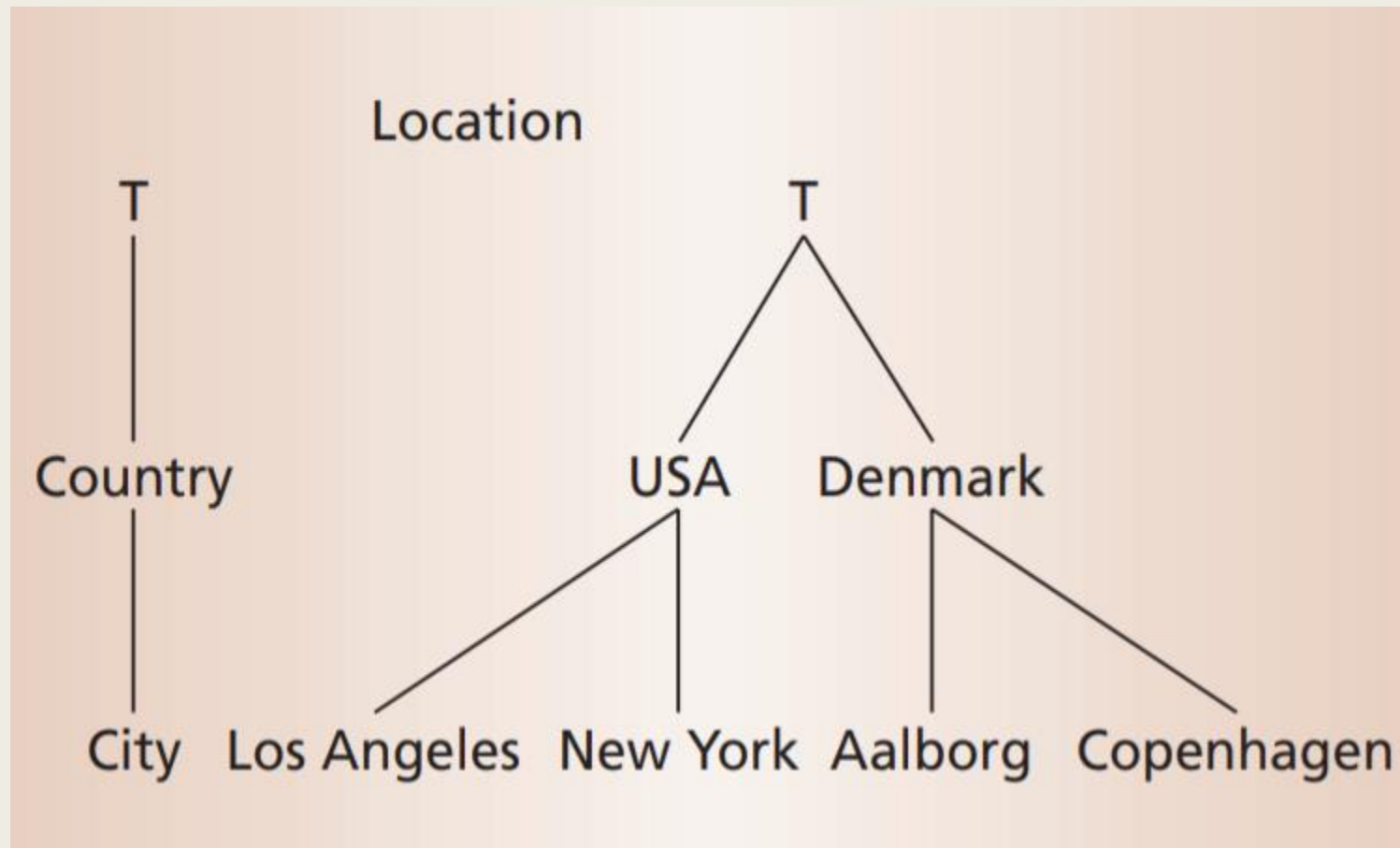
Куб с данни - фигура 1



Заявки (QUERYING)

- Заявките “Slice-and-dice” правят селекции за намаляване на куб. Например, можем да изрежем куба на Фигура 1, като разгледаме само тези клетки, които засягат хляба, а след това допълнително да намалим тези парчета, като разгледаме само клетките за 2000 година. Избирането на единична величина намалява размерите на куба.

Измерения на местоположения - фигура 2



- Заявките “Drill-down and roll-up” представляват две противоположни операции, които използват йерархията на измеренията и помага за осъществяването на агрегатни функции.

Заявки в дълбочина могат да бъдат прилагани върху схемата описана на Фигура 2.

- “Drill-across” заявките обединяват два или повече куба, които споделят едно или повече измерения. От гледна точка на релационните операции, можем да кажем, че се извършва обединение(join) между тях.
- Заявките “Rotating” позволяват на куба да бъде разглеждат под друг ъгъл с групирани данни в други измерение.

- Заявки “Ranking” могат да върнат само тези клетки, които се появяват в горната или долната част на посочената заявка - например 10-те най-продавани продукта в Копенхаген през 2000 г.

Представете по два примера за всеки тип складове за данни, които разгледахме:

Spreadsheets

Pivot table

Cube

Многоизмерни бази данни

ВЪПРОСИ ?

