

МОДЕЛИРАНЕ И АНАЛИЗ НА СОФТУЕР

Павел Кюркчиев

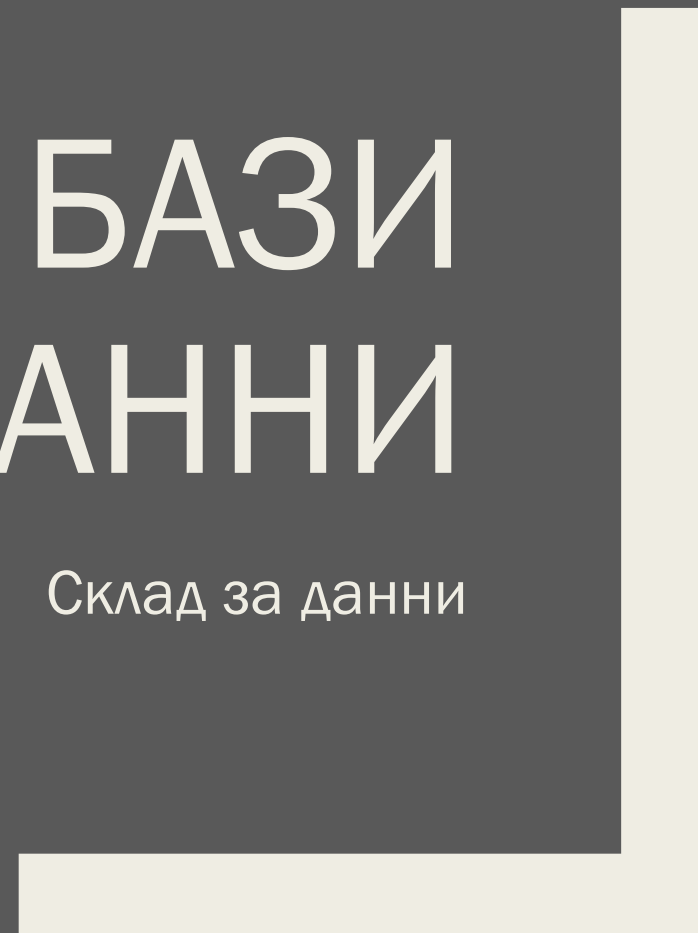
Ас. към ПУ „Паисий Хилендарски“

<https://github.com/pkyurkchiev>

@pkyurkchiev

МНОГОМЕРНИ БАЗИ ДАННИ

Склад за данни



Многомерна база данни

Multidimensional database (MDB)

- Многомерната база данни е специфичен тип база данни, оптимизирана за съхранение на данни и онлайн аналитична обработка (OLAP). Многомерната база данни е структура, комбинираща данни от различни източници, която осигурява връзки между базите данни и предлага мрежи, йерархии, масиви и други методи за форматиране на данните. В многомерна 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.

- Многомерните бази данни използват концепцията за куб данни (наричана още хиперкуб), за да представят измеренията на данните, които са достъпни за потребителите. Концепцията на многомерната 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 (производителност) в определени случаи в сравнение с традиционните релационни структури и бази данни. Този тип база данни обикновено е структурирана по начин, който оптимизира OLAP и приложения за съхранение на данни.

Типове Многомерните бази данни

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

Факты, Измерения (Величины), Мерки и Заявки

Факти (Facts)

- Фактите представляват обекта – модел на интерес или събитие в предприятието, което трябва да бъде анализирано, за да се разбере поведението му. В многомерните модели от данни фактите се дефинират имплицитно като комбинация на величините и техните стойности; Фактът съществува само ако има непразна клетка за определена комбинация от стойности.

- Факт е пример за конкретно събитие или явление 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 в или извън склад.

Образци (Snapshots)

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

Пример от реалния свят: Консервна кутия със зърна на рафт - 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 съществената и описателна част на многомерните бази данни. Важно в многомерните модели е да се използват измерения, които да осигурят възможно най-подробно описание (да предават значение) на фактите.

- Едно измерение е ключово описание (индекс), чрез което 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.

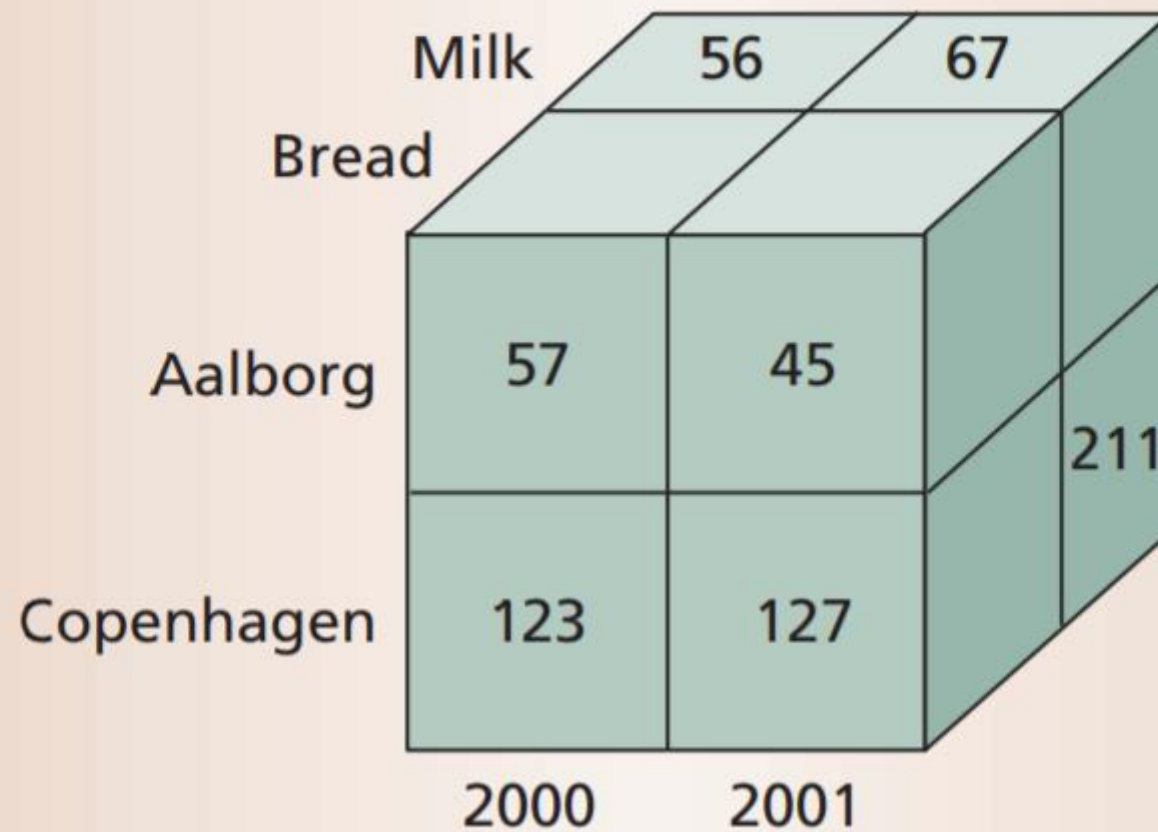
Мерки (Measures)

- В многомерна база данни мерките представляват свойствата на факта, които потребителят иска да оптимизира (обработи). Мярката е съвкупност от числови стойности на данни, като например сума, брой, минимум, максимум, средна стойност или персонализиран MDX израз, който създавате. Група от мерки представлява контейнер, съдържащ една или повече мерки.

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

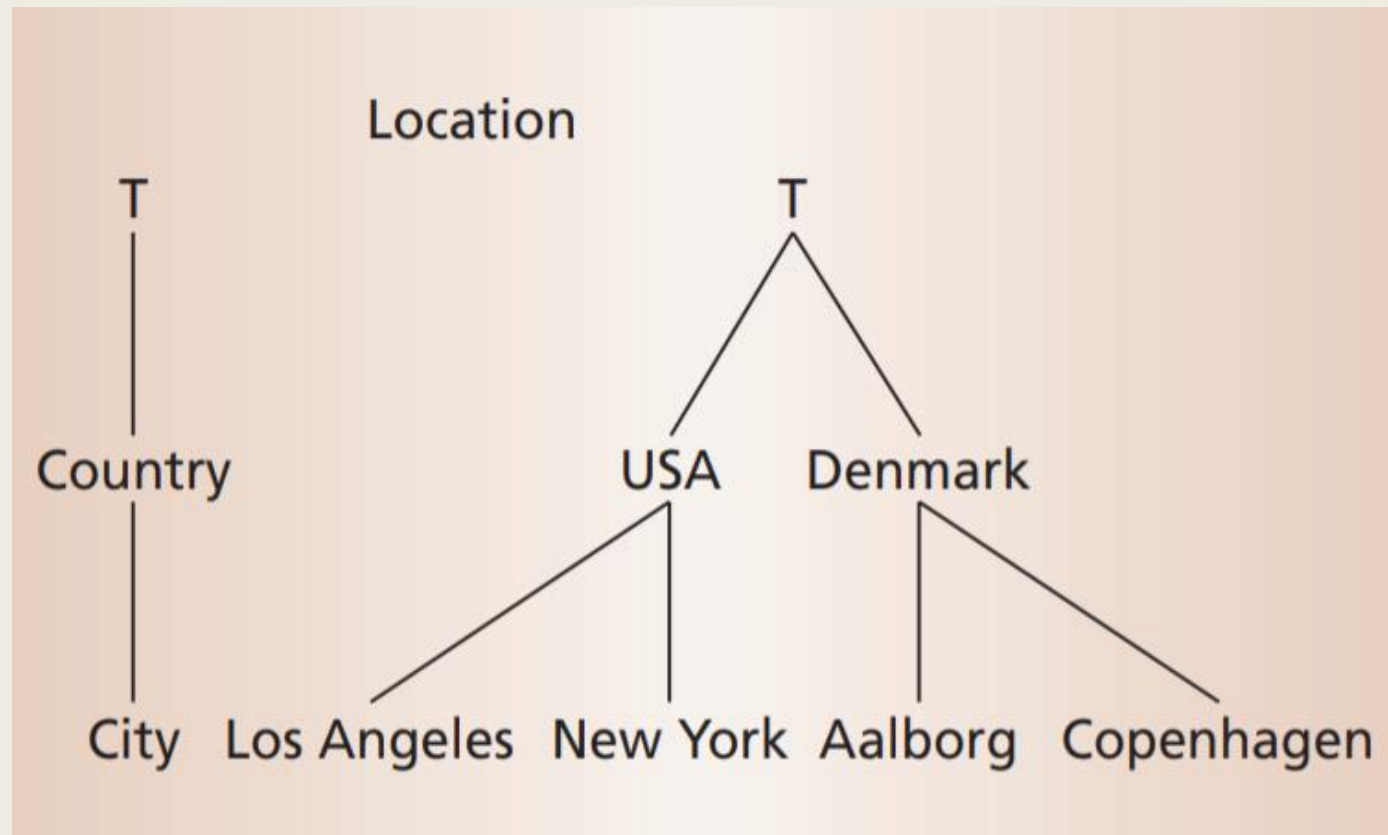


Заявки (QUERYING)

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

- 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 Фигурата



- „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 разглеждан под друг ъгъл с групирани данни в други измерения.
- 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

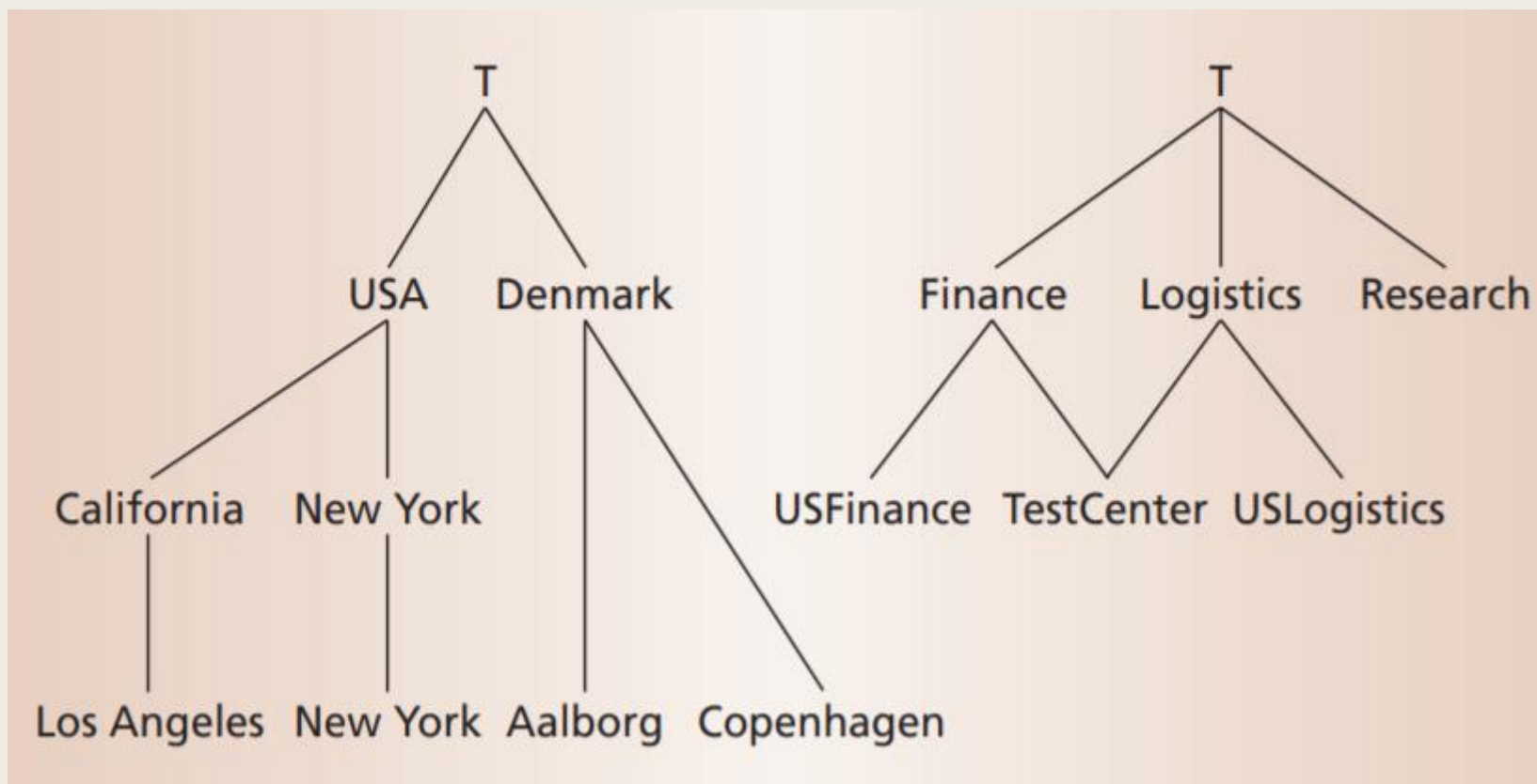
Създаване

- Multidimensional online analytical processing (MOLAP)
- Relational online analytical processing (ROLAP)

Multidimensional online analytical processing (MOLAP)

- MOLAP съхранява данни на диска в специализирани многомерни структури. Използва специализирани хеширания и индекси, за да извършва заявки върху данните.

Представяне - нерегулярни измерения



Relational online analytical processing (ROLAP)

- ROLAP използва специализирана релационна технология за съхранение на информация. За бързо извършване на заявките се използват специализирани индексни структури и различни bit-mapped индекси.
- Най – често имплементира схемите звезда или снежинка.

Представяне звезда

| ProductID | Product | Type | LocID | City | Country |
|-----------|---------|------|-------|---------|---------|
| 1 | Milk | Food | 1 | Aalborg | Denmark |

Product

Location

| ProductID | LocID | TimeID | Sale |
|-----------|-------|--------|------|
| 1 | 1 | 1 | 5.75 |

Sale (fact table)

| TimeID | Day | Month | Year |
|--------|-----|-------|------|
| 1 | 25 | May | 2001 |

Time

Представяне снежинка

| TypeID | Type |
|--------|------|
| 1 | Food |

Type

| CntID | Country |
|-------|---------|
| 1 | Denmark |

Country

| ProductID | Product | TypeID |
|-----------|---------|--------|
| 1 | Milk | 1 |

Product

| LocID | City | CntID |
|-------|---------|-------|
| 1 | Aalborg | 1 |

Location

| ProductID | LocID | TimeID | Sale |
|-----------|-------|--------|------|
| 1 | 1 | 1 | 5.75 |

Sale (fact table)

| TimeID | Day | MonthID |
|--------|-----|---------|
| 1 | 25 | 1 |

Day

| MonthID | Month | YearID |
|---------|-------|--------|
| 1 | May | 1 |

Month

| YearID | Year |
|--------|------|
| 1 | 2001 |

Year

Hybrid Online Analytical Processing (HOLAP)

- Използва комбинация за съхранение на данните от ROLAP и MOLAP. Като по този начин се прави комбинация между двата подхода.

ВЪПРОСИ ?

