

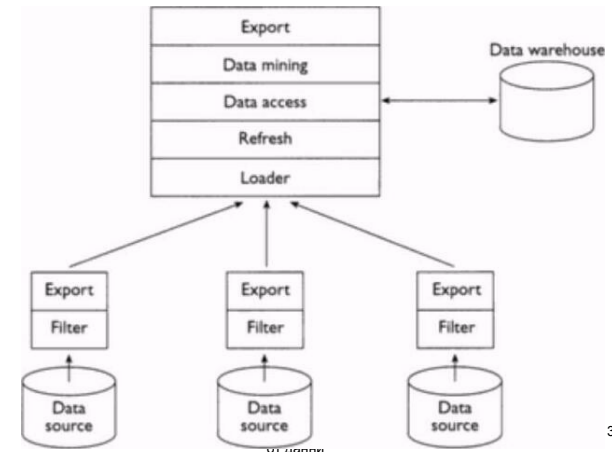
Складове с данни

Бази за анализ

Богдан Шишеджиев - Складове
от данни

1

Архитектура



Предназначение

- Типове обработки
 - OLTP (*On Line Transaction Processing*) - транзакционни
 - OLAP (*On-line Analytical Processing*) - аналитични
- Типове бази
 - транзакционни –
 - голям брой потребители
 - нестационарни (примигващи)
 - отразяващи текущото състояние
 - критични (силно натоварени)
 - складове
 - малък брой потребители (анализатори)
 - сравнително стационарни
 - отразяват историята на данните (натрупване на текущи състояния)
 - слабо натоварени

Богдан Шишеджиев - Складове
от данни

2

Архитектура

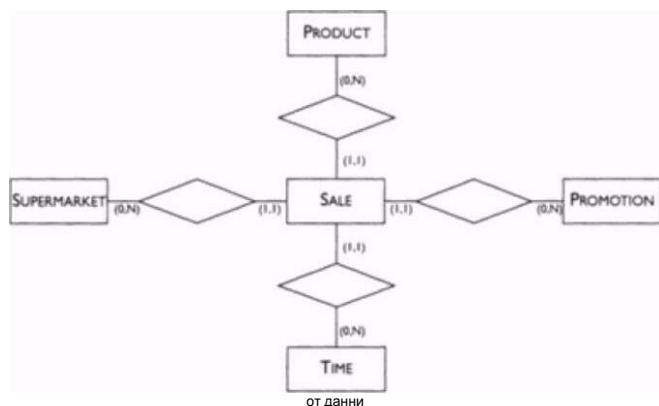
- Компоненти на източника
 - филтър – отделя и проверява кохерентността на данните за експорт
 - експорт – извършва зареждане на определени порции във времето.
- Компоненти на склада
 - лодер – начално зареждане и подготовка на склада за работа.
 - обновяване – зарежда порциите
 - достъп
 - data mining
 - експорт – към други складове. Създава се йерархия от складове

Богдан Шишеджиев - Складове
от данни

4

Схеми за складове

- Звезда

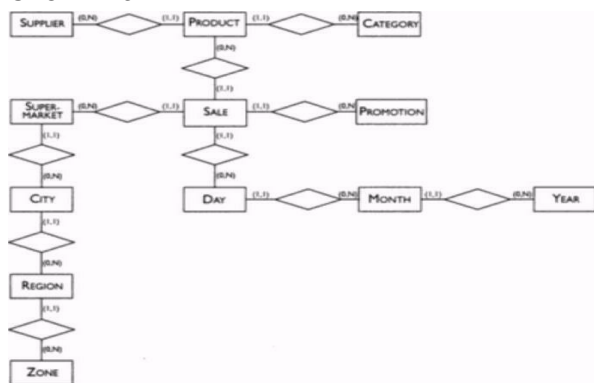


Пример – Верига супермаркети



Схеми за складове

- Снежинка



Пример – Верига супермаркети

- Таблица на фактите – в BCNF
 - измервани стойности
 - числови
 - адитивни
 - полуадитивни
 - неадитивни
 - текстови
 - детайлност на фактите – всяка продажба на продукт
- Дименсии – ненормализирани, с голям излишък
 - атрибути - текстови

Пример – Верига супермаркети

- Изчисление на нужното дисково пространство
 - Dimension time : 2 год. x 365 дни = 730 дни
 - Dimension shop : 300 магазина, запис на продажбите всеки ден
 - Dimension product : 30.000 продукта във всеки магазин, като 3000 се продават всеки ден в даден магазин.
 - Dimension promotion : един артикул може да участва в един ден и в един магазин само в една промоция
 - Записи на елементарни факти $300 \times 730 \times 3000 \times 1 = 657 \cdot 10^6$ записа
 - Брой ключови полета 4; брой полета със стойности 4 ; общ брой полета =8
 - Размер на таблицата на фактите - $657 \cdot 10^6 \times 8$ полета $\times 4B = 21 GB$

Богдан Шишеджиев - Складове
от данни

9

Операции за анализ на данните

• Пример

```
select Time.Month, Product.Name, sum(Qty)
from Sale, Time, Product, Promotion
where Sale.TimeCode = Time.TimeCode
and Sale.ProductCode = Product.ProductCode
and Sale.PromoCode = Promotion.PromoCode
and (Product.Name = 'Pasta' or Product.Name = 'Oil')
and Time.Month between 'Feb' and 'Apr'
and Promotion.Name = 'SuperSaver'
group by Time.Month, Product.Name
order by Time.Month, Product.Name
Pivot Time.Month
```

	Feb	Mar	Apr
Oil	5K	5K	7K
Pasta	45K	50K	51K

Богдан Шишеджиев - Складове
от данни

11

Операции за анализ на данните

• Обща форма на оператор SQL

```
select D1.C1, ... Dn.Cn, Aggr1(F,C1),..., Aggrn(F,Cn)
from Fact as F, Dimension1 as D1,... DimensionN as Dn
where join-condition (F, D1)
and...
and join-condition (F, Dn)
and selection-condition
group by D1.C1, ... Dn.Cn
order by D1.C1, ... Dn.C
```

Богдан Шишеджиев - Складове
от данни

10

Операции drill-down и roll-up

• drill-down – добавяне на дименсия по-която се агрегат резултатите

Time.Month	Product.Name	sum(Qty)
Feb	Pasta	45K
Mar	Pasta	50K
Apr	Pasta	51K

Time.Month	Product.Name	Zone	sum(Qty)
Feb	Pasta	North	18K
Feb	Pasta	Centre	18K
Feb	Pasta	South	12K
Mar	Pasta	North	18K
Mar	Pasta	Centre	18K
Mar	Pasta	South	14K
Apr	Pasta	North	18K
Apr	Pasta	Centre	17K
Apr	Pasta	South	16K

12

Операции drill-down и roll-up

- roll-up – премахване на дименсия по-която се агрегират резултатите

Time.Month	Product.Name	Zone	sum(Qty)
Feb	Pasta	North	18K
Feb	Pasta	Centre	18K
Feb	Pasta	South	12K
Mar	Pasta	North	18K
Mar	Pasta	Centre	18K
Mar	Pasta	South	14K
Apr	Pasta	North	18K
Apr	Pasta	Centre	17K
Apr	Pasta	South	16K

Product.Name	Zone	sum(Qty)
Pasta	North	54K
Pasta	Centre	50K
Pasta	South	42K

13

Куб данни

Make	Year	Colour	sum(Sales)
Ferrari	1998	Red	50
Ferrari	1999	Red	85
Ferrari	1998	ALL	50
Ferrari	1999	ALL	85
Ferrari	ALL	Red	135
Ferrari	ALL	ALL	135
Porsche	1998	Red	80
Porsche	1998	ALL	80
Porsche	ALL	Red	80
Porsche	ALL	ALL	80
ALL	1998	Red	130
ALL	1999	Red	85
ALL	ALL	Red	215
ALL	1998	ALL	130
ALL	1999	ALL	85
ALL	ALL	ALL	215

Богдан Шишеджиев - Складове
от данни

15

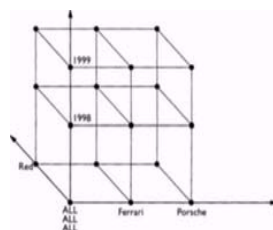
Куб данни

- Клауза **with cube**
- ```

select Make, Year, Colour, sum(Sales)
from Sales
where (Make = 'Ferrari' or Make =
'Porsche')
and Colour = 'Red'
and Year between 1998 and 1999
group by Make, Year, Colour
with cube

```

| Make    | Year | Colour | Sales |
|---------|------|--------|-------|
| Ferrari | 1998 | Red    | 50    |
| Ferrari | 1999 | Red    | 85    |
| Porsche | 1998 | Red    | 80    |



Богдан Шишеджиев - Складове  
от данни

14