OLAP

joost.vennekens@kuleuven.be

Toepassingen







Waarom?

- Trouwe klanten belonen
- Gegevens verzamelen
  - Facebook model

Waarom?

- Grote databank
  - Produkten
  - Produkten - winkels
  - Produkten - produkten
  - Klanten
  - Klanten - winkels
  - Klanten - produkten

How Target knows when its shoppers are pregnant - and figured out a teen was before her father did

By NINA GOLGOWSKI
UPDATED: 13:19 GMT, 18 February 2012

Share

19 View comments

Hearing that stores like Target monitor their customers' spending habits to determine their future ones may not be much of a shock to most.

What may be more of one is that Target's department for Guest Marketing Analytics is so in tune with their consumers based on their spending, they can predict major changes in their lives.

They in fact do it so well, as the New York Times reports, that in one case they were able to predict a high school girl's pregnancy before her father did.

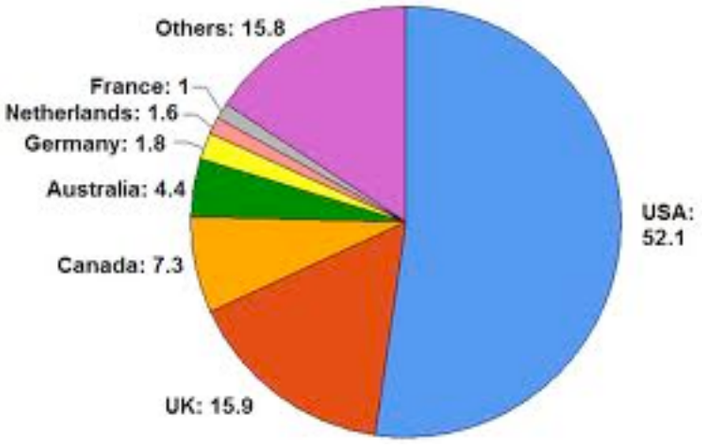


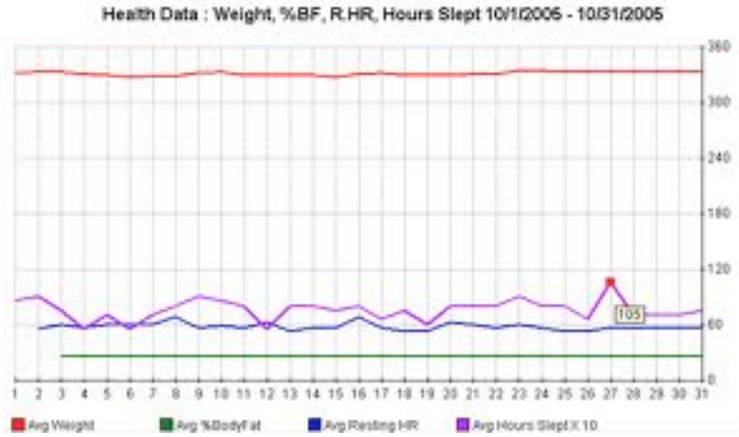
One job: Championing consumer predictions, Target tracks their consumers' spending habits using ID numbers for each shopper which are given to their very own Guest Marketing Analystics department

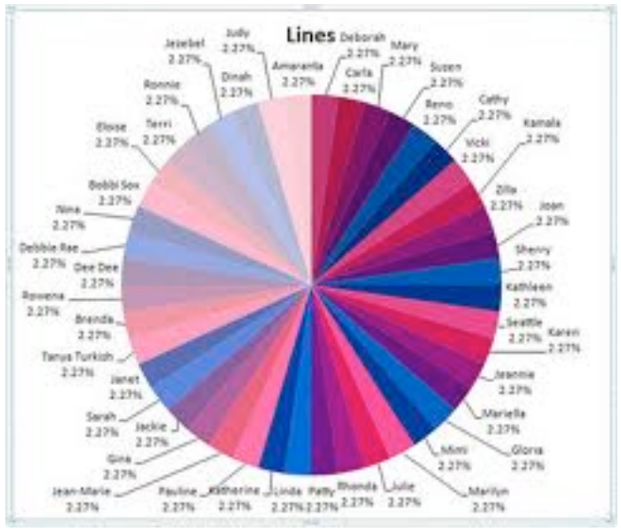
Meeting with a Target manager while clutching some printed ads that had been delivered to his home, a customer was outraged by their product placement aimed at his young daughter.

'My daughter got this in the mail!' the man said to the manager according to the Times, showing him the store's coupons advertising baby-related products.

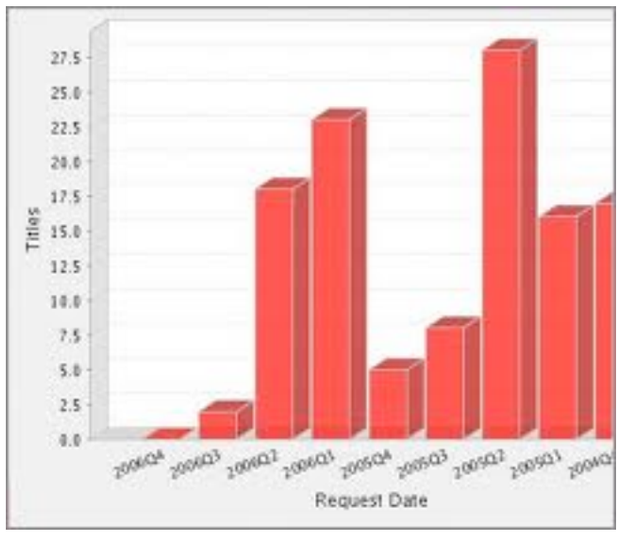
Doel

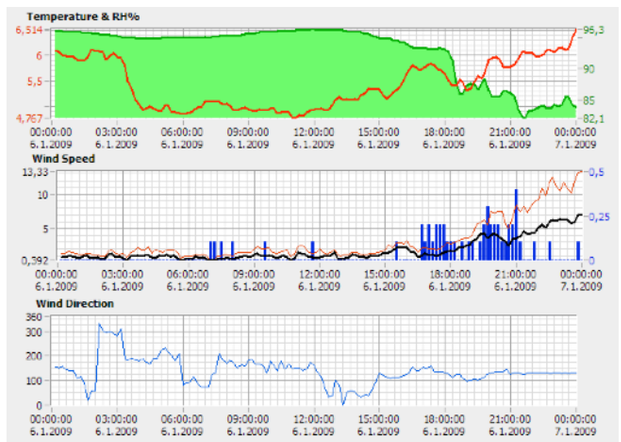


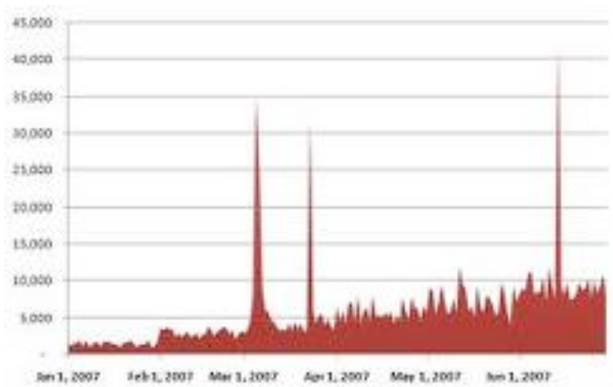




- Interessante patronen
- Uit hele hoop gedetailleerde data







OLAP vs. OLTP

- Online Transaction Processing: dagdagelijks
  - Veel kleine bewerkingen op stukjes van DB
  - Hoge gebruiksgraad -> performant (latency)
  - Moet up-to-date zijn
- Online Analytical Processing
  - Paar grote bewerkingen op hele DB
  - Periodiek gevoed vanuit OLTP DB

Termen

- Business Intelligence (BI): globale patronen die uit data gehaald worden
- Data warehouse: OLAP databank
- Data mart: date warehouse dat gefocust is op een bepaald aspect
- Data mining: kennis genereren uit data

Termen

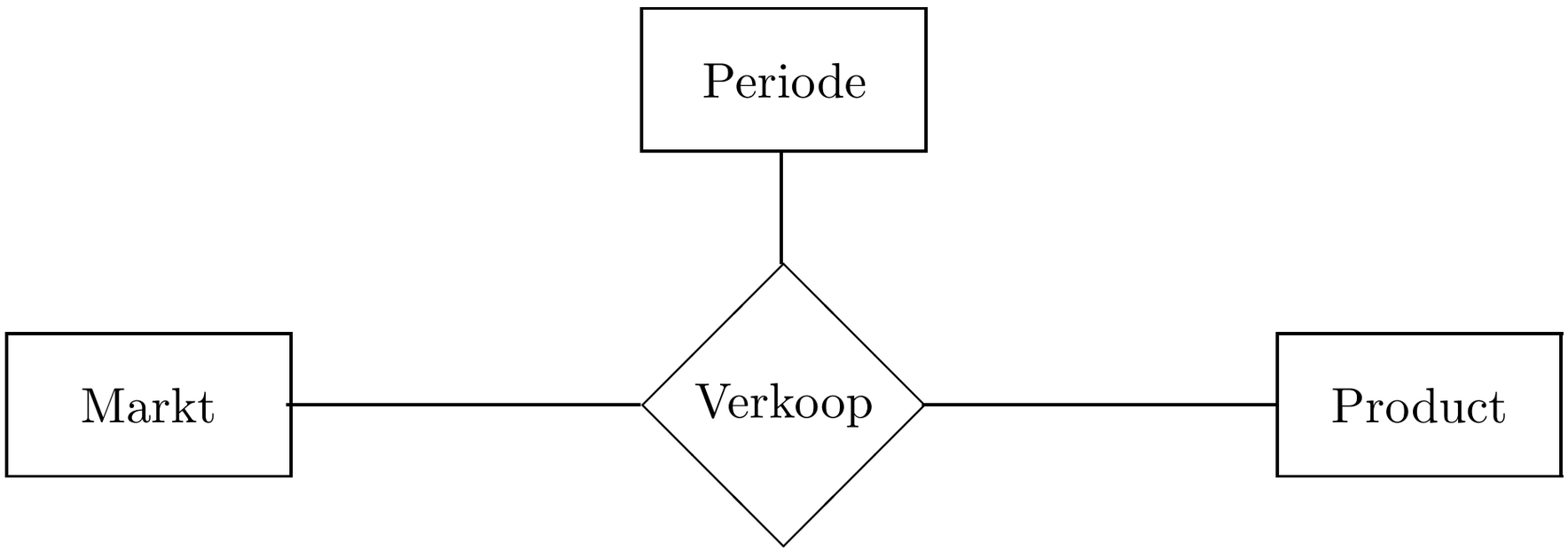
- Data mining: kennis genereren uit data
  - Associatie: bier + chips
  - Classificatie: dit soort klanten kopen product X
  - Clustering: verschillende klantenprofielen

Hoe zien OLAP data eruit?

Bijvoorbeeld

- Supermarkt verkoopt
  - verschillende produkten
  - in verschillende vestigingen
  - in verschillende periodes

Ster schema



Tabellen

Feitentabel

Verkoop

markt-id	pro-id	tim-id	hoev
M1	P1	T1	1000
M1	P2	T1	2000
M1	P3	T1	1500
M1	P4	T1	2500
M2	P1	T1	500
M2	P2	T1	800
M2	P3	T1	0
M2	P4	T1	3333
M3	P1	T1	5000
M3	P2	T1	8000
M3	P3	T1	10
M3	P4	T1	3300
M1	P1	T2	1001
M1	P2	T2	2001
M1	P3	T2	1501
M1	P4	T2	2501
M2	P1	T2	501
M2	P2	T2	801
M2	P3	T2	1

Dimensie tabellen

Markt

markt-id	stad	prov	gewest
M1	Lier	Antwerpen	Vlaanderen
M2	Tongeren	Limburg	Vlaanderen
M3	Spa	Luik	Wallonië

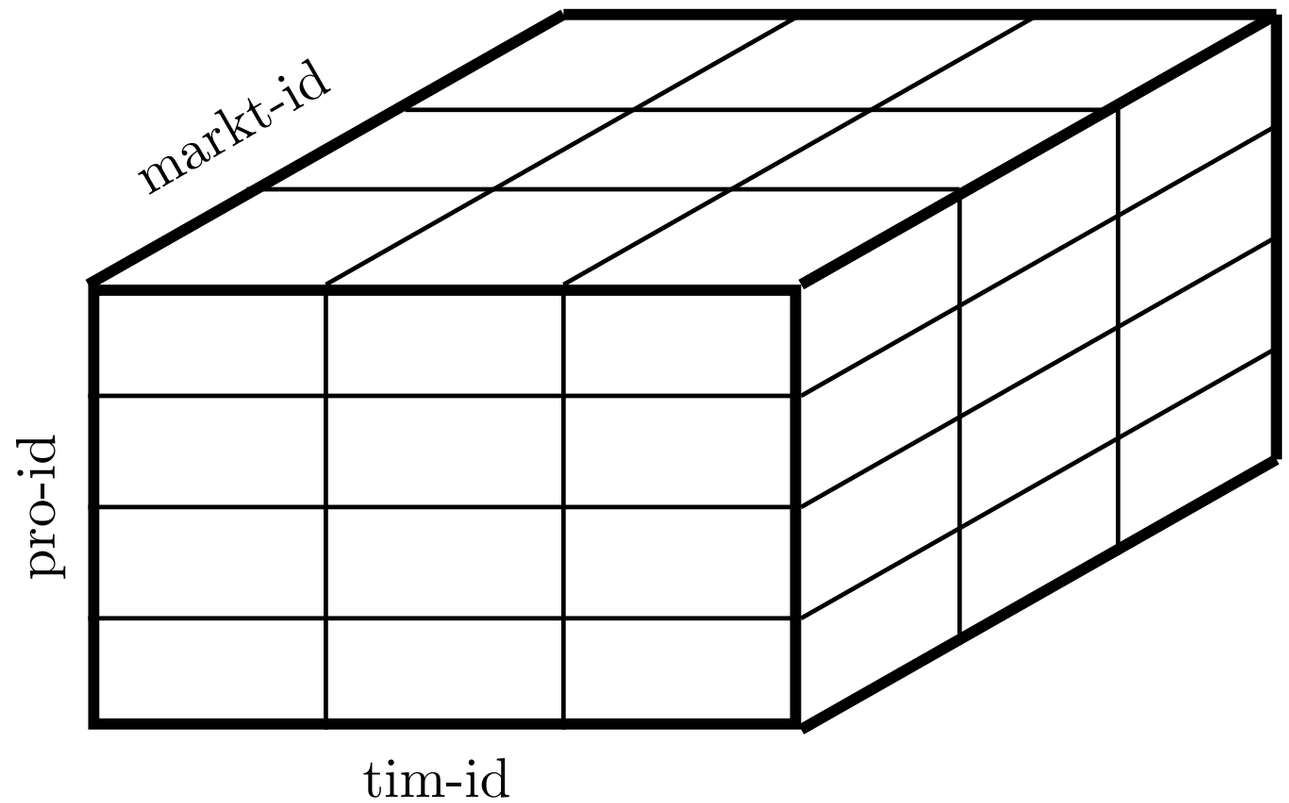
Product

pro-id	naam	soort	prijs
P1	bier	drank	1.10
P2	zakdoekjes	zachtgoed	2.70
P3	hesp	vlees	3.90
P4	frisdrank	drank	1.05

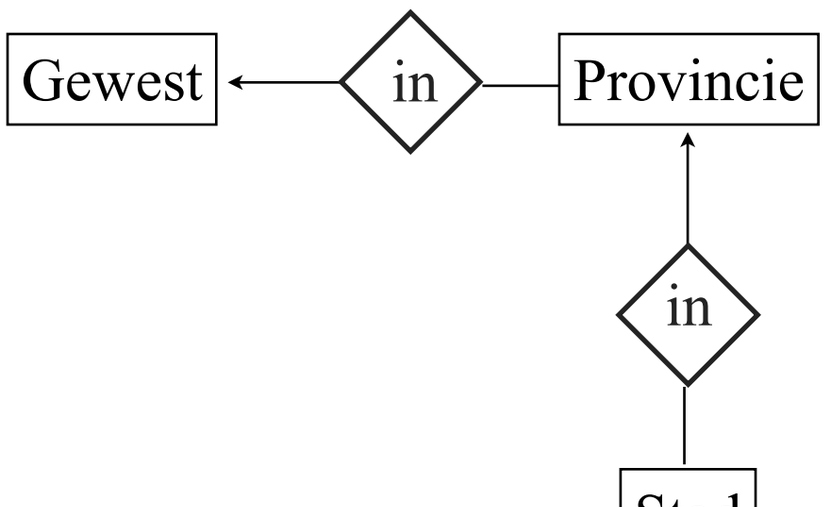
Periode

tim-id	week	maand	kwartaal
T1	1	januari	1
T2	23	juni	2
T3	51	december	4

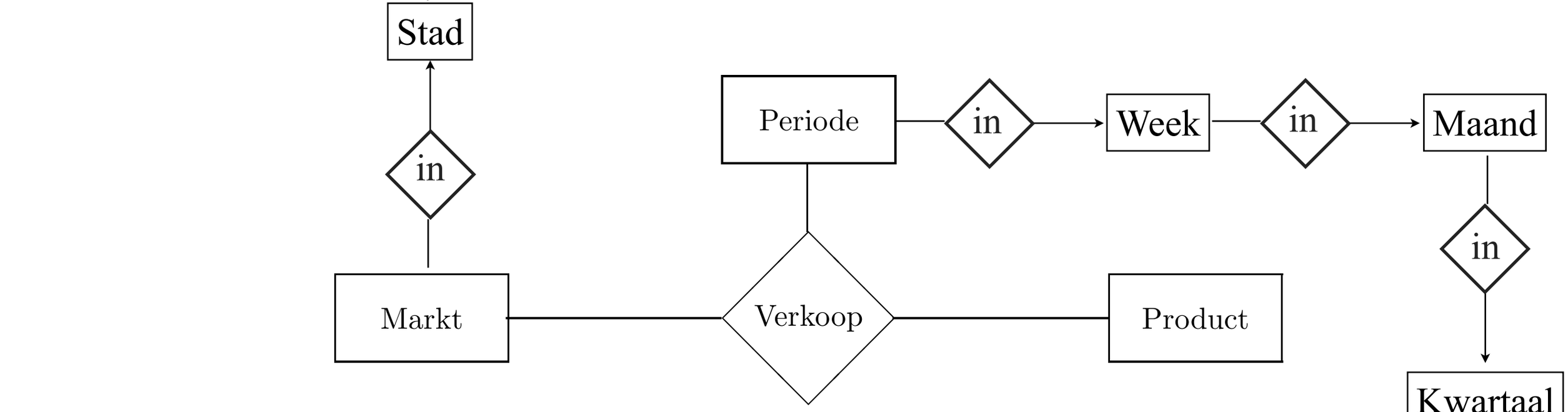
Data kubus



Sneeuwvlok Schema

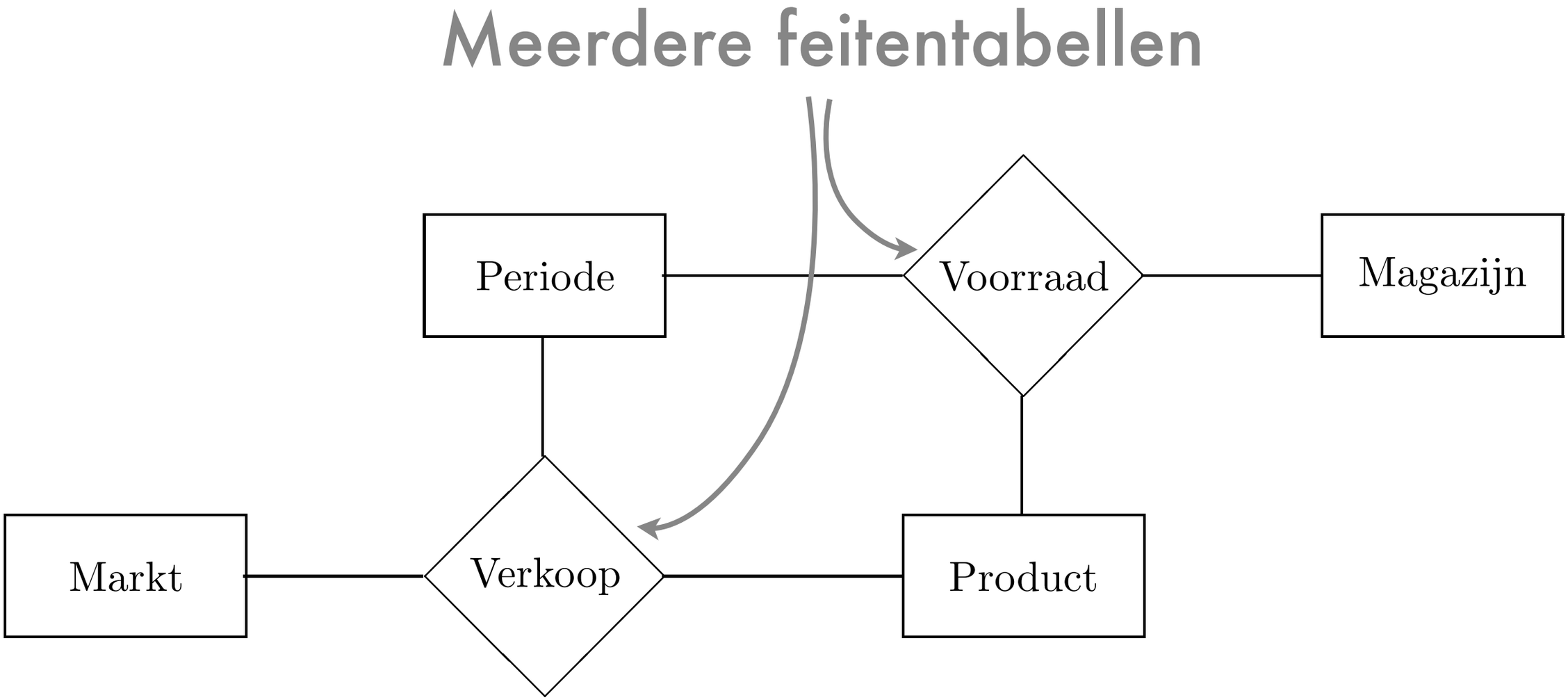


Periode			
tim-id	week	maand	kwartaal
T1	1	januari	1
T2	23	juni	2
T3	51	december	4



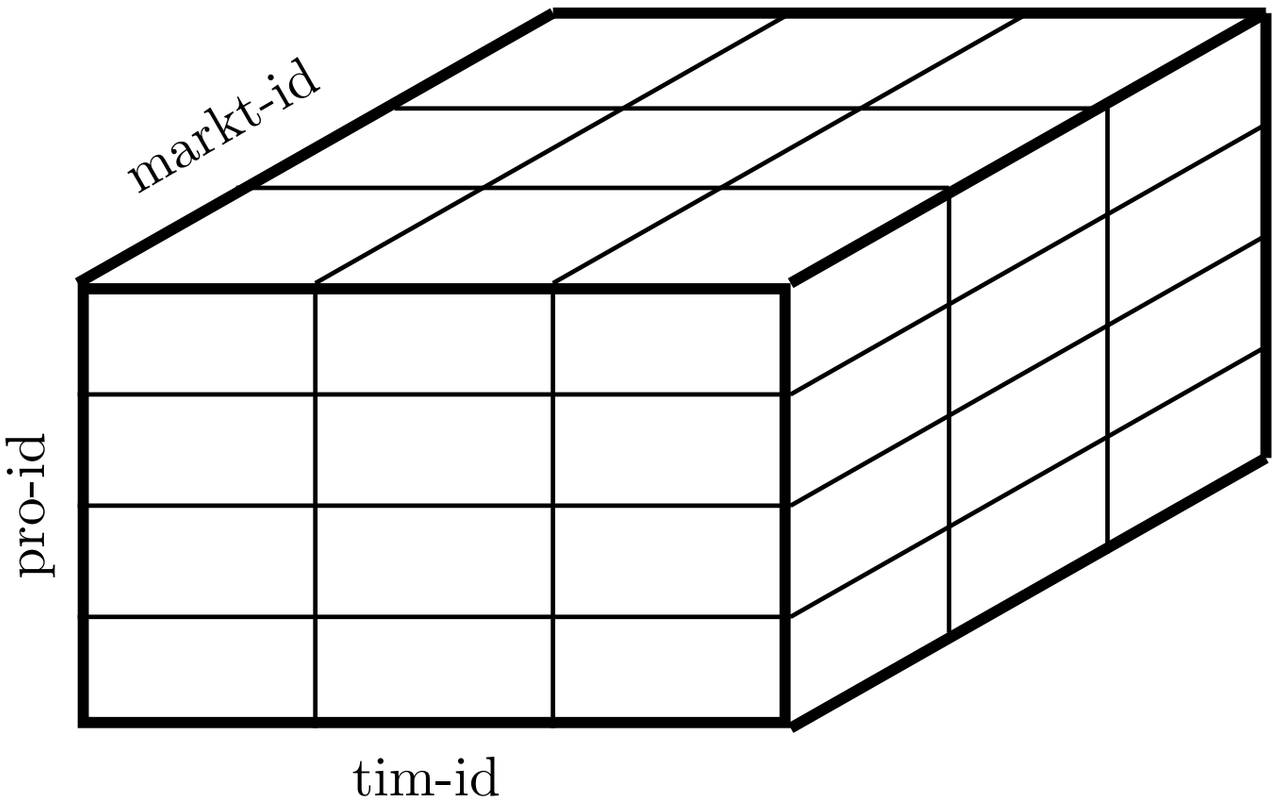
Markt			
markt-id	stad	prov	gewest
M1	Lier	Antwerpen	Vlaanderen
M2	Tongeren	Limburg	Vlaanderen
M3	Spa	Luik	Wallonië

Constellatie Schema



Wat doen we ermee?

Aggregatie



Weken optellen

Reductie: 3D => 2D

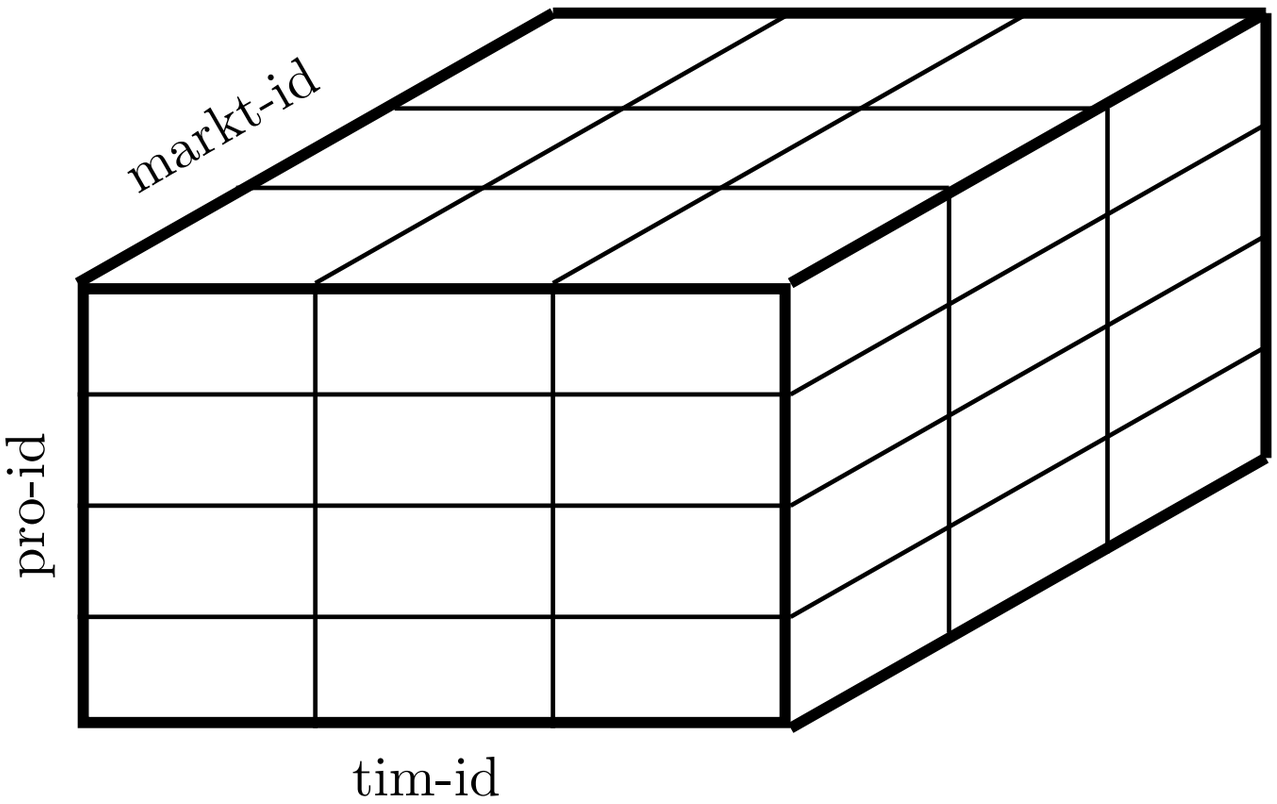
Produkten optellen

Reductie: 2D => 1D

totqty		markt-id			totaal
		M1	M2	M3	
pro-id	P1	3003	1502	15003	19509
	P2	6003	2403	24003	32409
	P3	4503	3	33	4539
	P4	7503	7000	9903	24406
totaal		21012	10909	48942	80863

Reductie: 1D => OD

Aggregatie

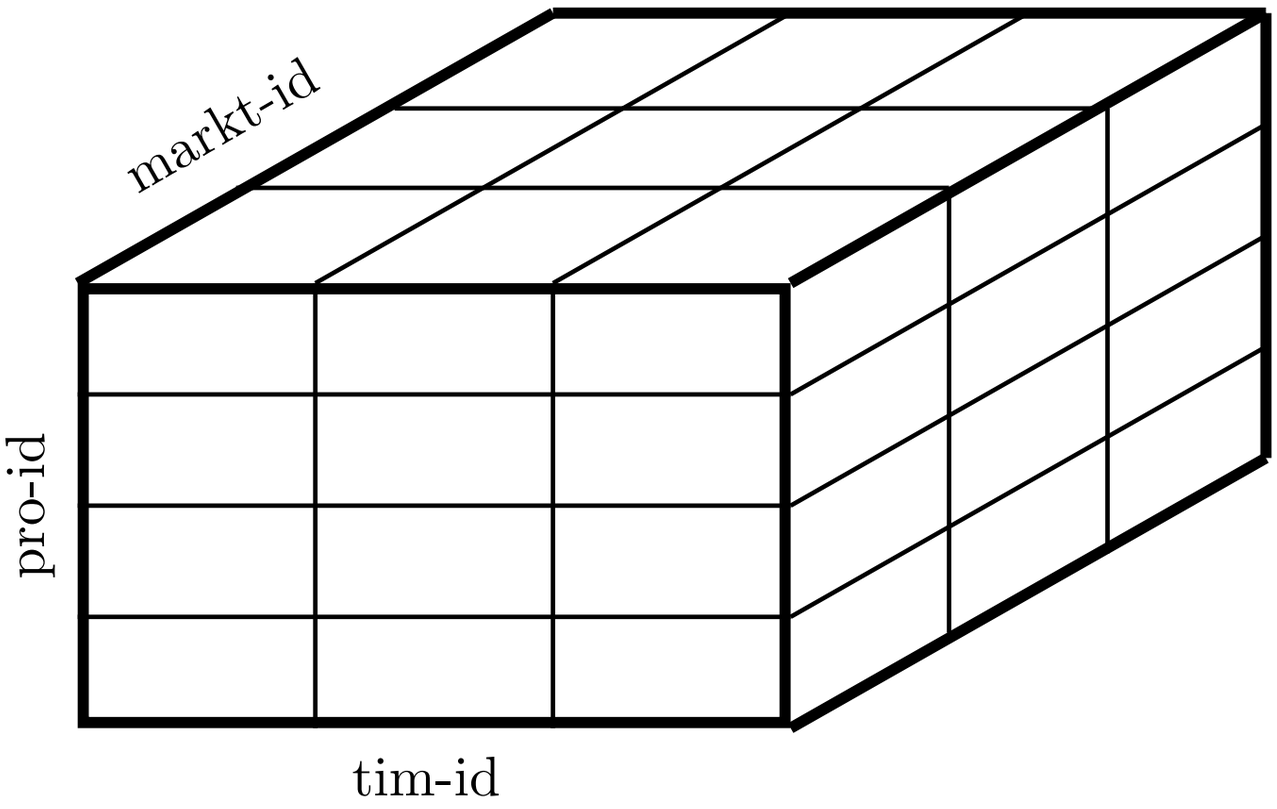


alles optellen

select sum(hoev)
from Verkoop

totqty		markt-id			
		M1	M2	M3	totaal
pro-id	P1	3003	1502	15003	19509
	P2	6003	2403	24003	32409
	P3	4503	3	33	4539
	P4	7503	7000	9903	24406
totaal		21012	10909	48942	80863

Aggregatie

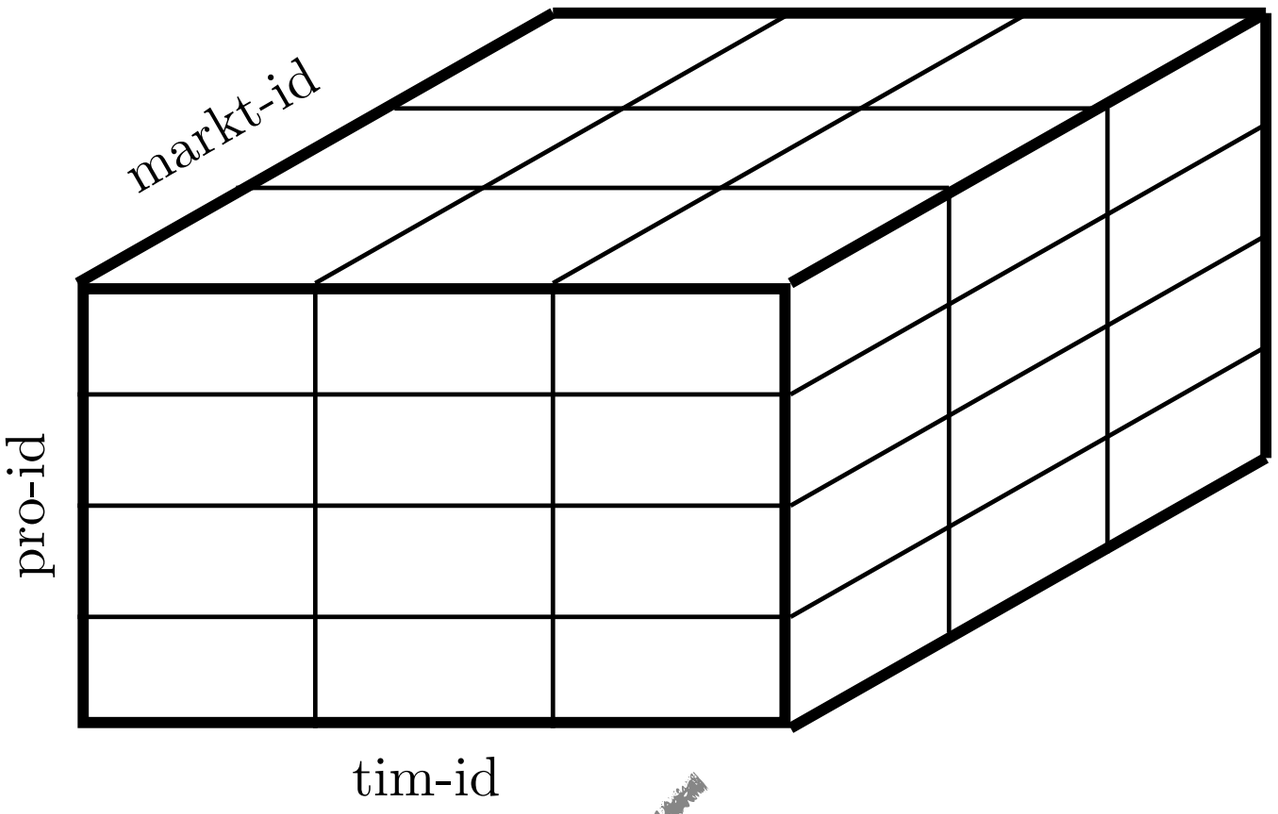


Weken
markten
optellen

select pro-id,
sum(hoev)
from Verkoop
group by pro-id

totqty		markt-id			
		M1	M2	M3	totaal
pro-id	P1	3003	1502	15003	19509
	P2	6003	2403	24003	32409
	P3	4503	3	33	4539
	P4	7503	7000	9903	24406
totaal		21012	10909	48942	80863

Aggregatie

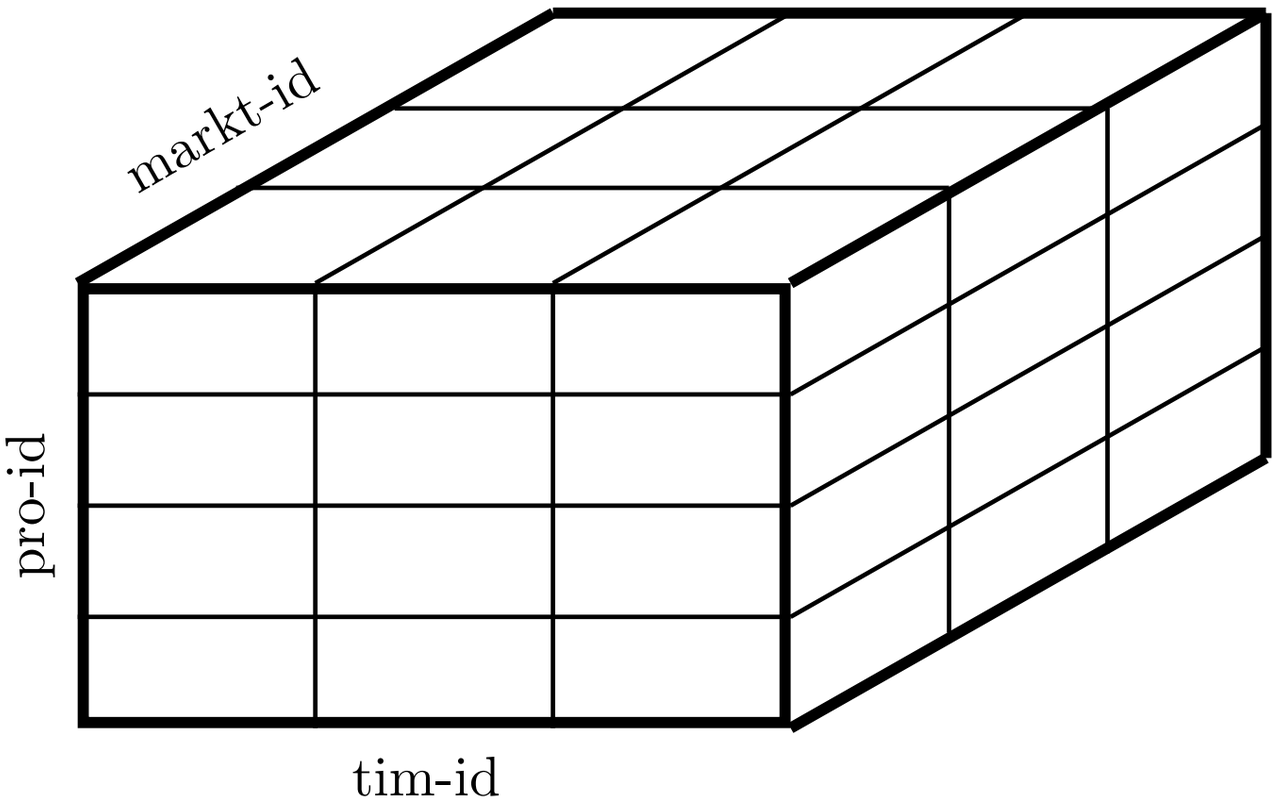


select markt-id, sum(hoev)
from Verkoop
group by markt-id;

Weken
produkten
optellen

totqty		markt-id			
		M1	M2	M3	totaal
pro-id	P1	3003	1502	15003	19509
	P2	6003	2403	24003	32409
	P3	4503	3	33	4539
	P4	7503	7000	9903	24406
totaal		21012	10909	48942	80863

Met SQL



Weken optellen

select pro-id, markt-id,
sum(hoev)
from Verkoop
group by prod-id, markt-id

totqty		markt-id		
		M1	M2	M3
	P1	3003	1502	15003
pro-id	P2	6003	2403	24003
	P3	4503	3	33
	P4	7503	7000	9903

Met SQL

- **select** pro-id, markt-id, sum(hoev)
  from Verkoop **group by** prod-id, markt-id;

- **select** pro-id, sum(hoev)
  from Verkoop **group by** prod-id;

- **select** markt-id, sum(hoev)
  from Verkoop **group by** markt-id;

- **select** sum(hoev) as totqty
  from Verkoop;

Veel werk voor
gebruiker
én systeem

Grouping sets

- **select** markt-id, pro-id, sum(hoev)
  from Verkoop **group by**
  grouping sets ((martk-id), (pro-id));

markt-id	pro-id	totqty
M1	*null*	21012
M2	*null*	10909
M3	*null*	48942
null	P1	19509
null	P2	32409
null	P3	4539
null	P4	24406

- Is geen relatie: *null* betekent “geen deel van groeping set”

totqty		markt-id			
		M1	M2	M3	totaal
pro-id	P1	3003	1502	15003	19509
	P2	6003	2403	24003	32409
	P3	4503	3	33	4539
	P4	7503	7000	9903	24406
totaal		21012	10909	48942	80863

- **select** markt-id, pro-id, sum(hoev)
  from Verkoop **group by**
  grouping sets ((markt-id, prod-id),
  (martk-id), (pro-id),
  ());

Rollup

- Combinatie van grouping sets
- **select** markt-id, pro-id, sum(hoev)
  from Verkoop **group by**
  rollup (markt-id, pro-id);
- grouping sets
  ((markt-id, prod-id),
    (markt-id),
    ())

markt-id	pro-id	totqty
M1	P1	3003
M1	P2	6003
M1	P3	4503
M1	P4	7503
M2	P1	1503
M2	P2	2403
M2	P3	3
M2	P4	7000
M3	P1	15003
M3	P2	24003
M3	P3	33
M3	P4	9903
M1	*null*	21012
M2	*null*	10909
M3	*null*	48942
null	*null*	80863

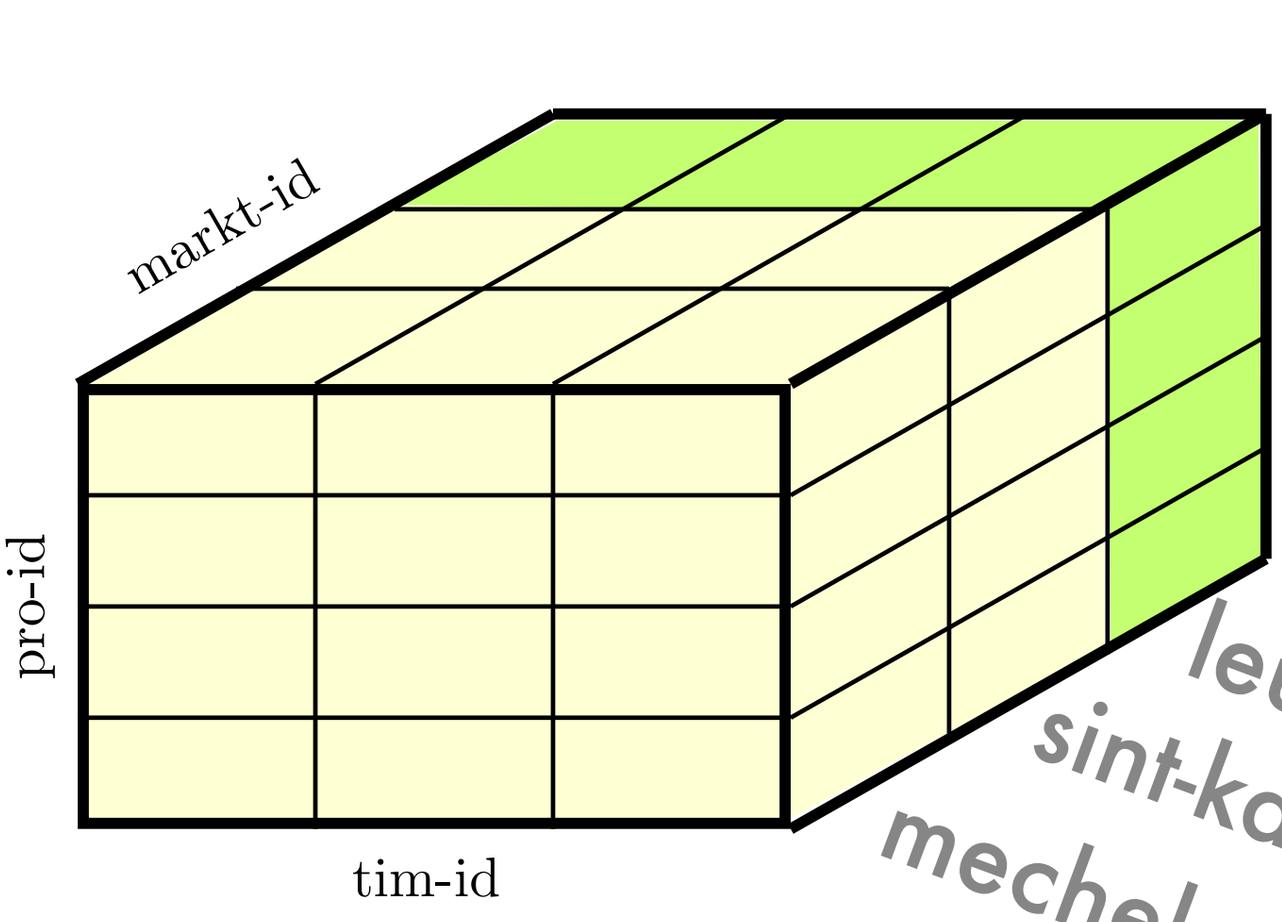
Cube

- Andere combinatie van grouping sets
- **select** markt-id, pro-id, sum(hoev)
  from Verkoop **group by**
  cube (markt-id, pro-id);
- grouping sets
  ((markt-id, prod-id),
  (markt-id), (prod-id),
  ())

markt-id	pro-id	totqty
M1	P1	3003
M1	P2	6003
M1	P3	4503
M1	P4	7503
M2	P1	1503
M2	P2	2403
M2	P3	3
M2	P4	7000
M3	P1	15003
M3	P2	24003
M3	P3	33
M3	P4	9903
M1	*null*	21012
M2	*null*	10909
M3	*null*	48942
null	P1	19509
null	P2	32409
null	P3	4539
null	P4	24406
null	*null*	80863

Fijnere aggregatie

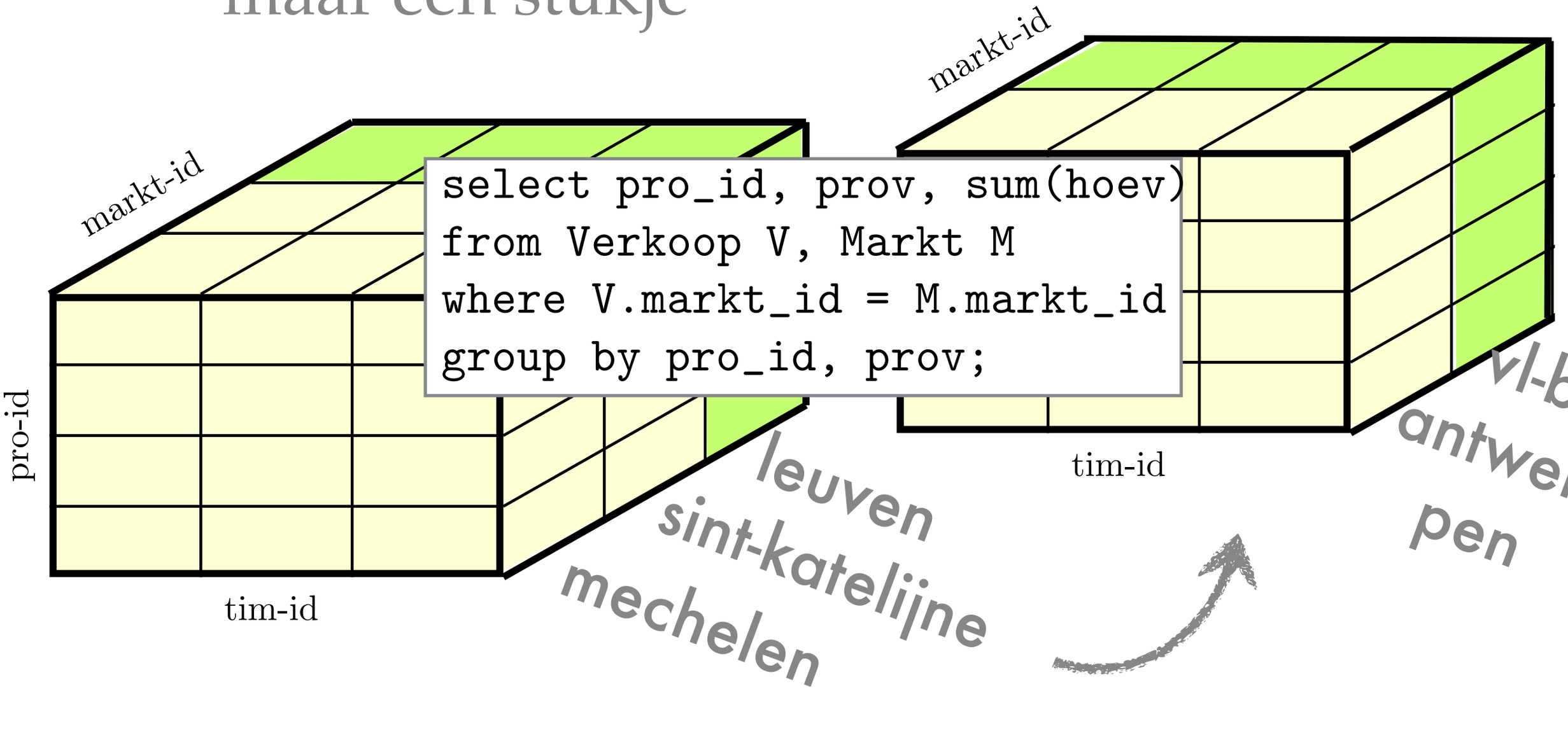
- Ipv. hele dimensie wegwerken, enkel maar een stukje





Fijnere aggregatie

- Ipv. hele dimensie wegwerken, enkel maar een stukje



Rolling up

markt-id → stad → prov → gewest

```
select pro_id, gewest, sum(hoev)
from Markt m, Prov-verkoop p
where p.prov = m.prov
group by pro_id, gewest;
```

Rolling up

	Vlaanderen	Wallonië
P1	4506	15003
P2	8406	24003
P3	4506	33
P4	14503	9903

Herbruik mogelijk

```
select pro_id, prov, sum(hoev)
from Verkoop V, Markt M
where V.markt_id = M.markt_id
group by pro_id, prov;
```

	Antwerpen	Limburg	Luik
P1	3003	1503	15003
P2	6003	2403	24003
P3	4503	3	33
P4	7503	7000	9903

Drilling down vs. rolling up

markt-id $\rightarrow$ stad $\rightarrow$ prov $\rightarrow$ gewest

```
select pro_id, gewest, sum(hoev)
from Verkoop V, Markt M
where V.markt_id = M.markt_id
group by pro_id, gewest;
```

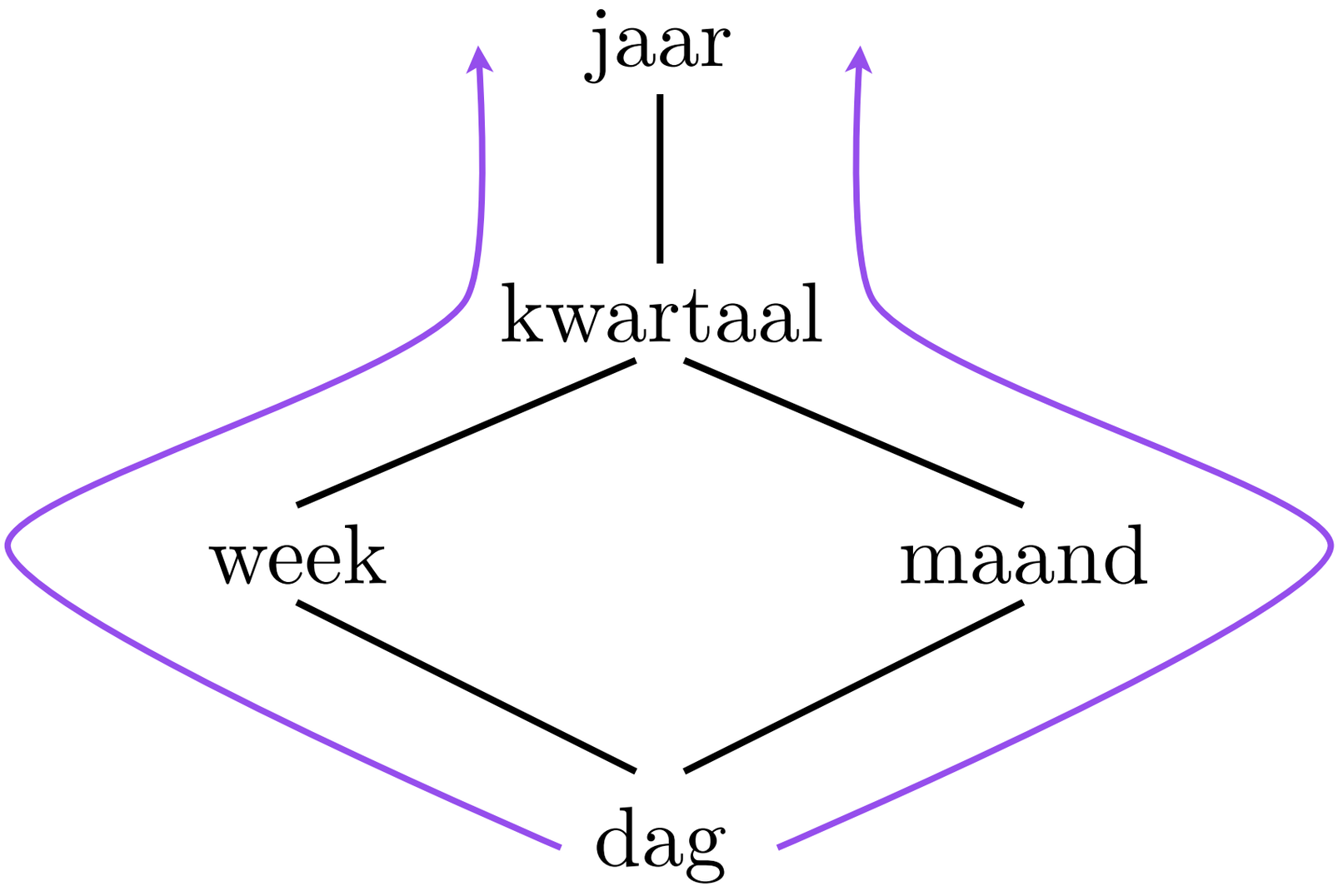
```
select pro_id, prov, sum(hoev)
from Verkoop V, Markt M
where V.markt_id = M.markt_id
group by pro_id, prov;
```

Drilling down

	Vlaanderen	Wallonië
P1	4506	15003
P2	8406	24003
P3	4506	33
P4	14503	9903

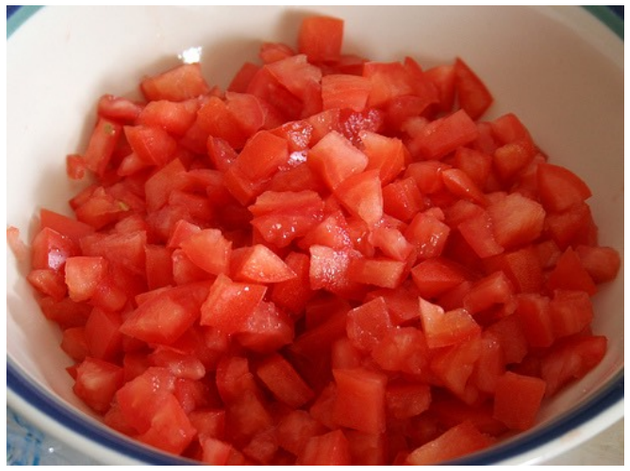
	Antwerpen	Limburg	Luik
P1	3003	1503	15003
P2	6003	2403	24003
P3	4503	3	33
P4	7503	7000	9903

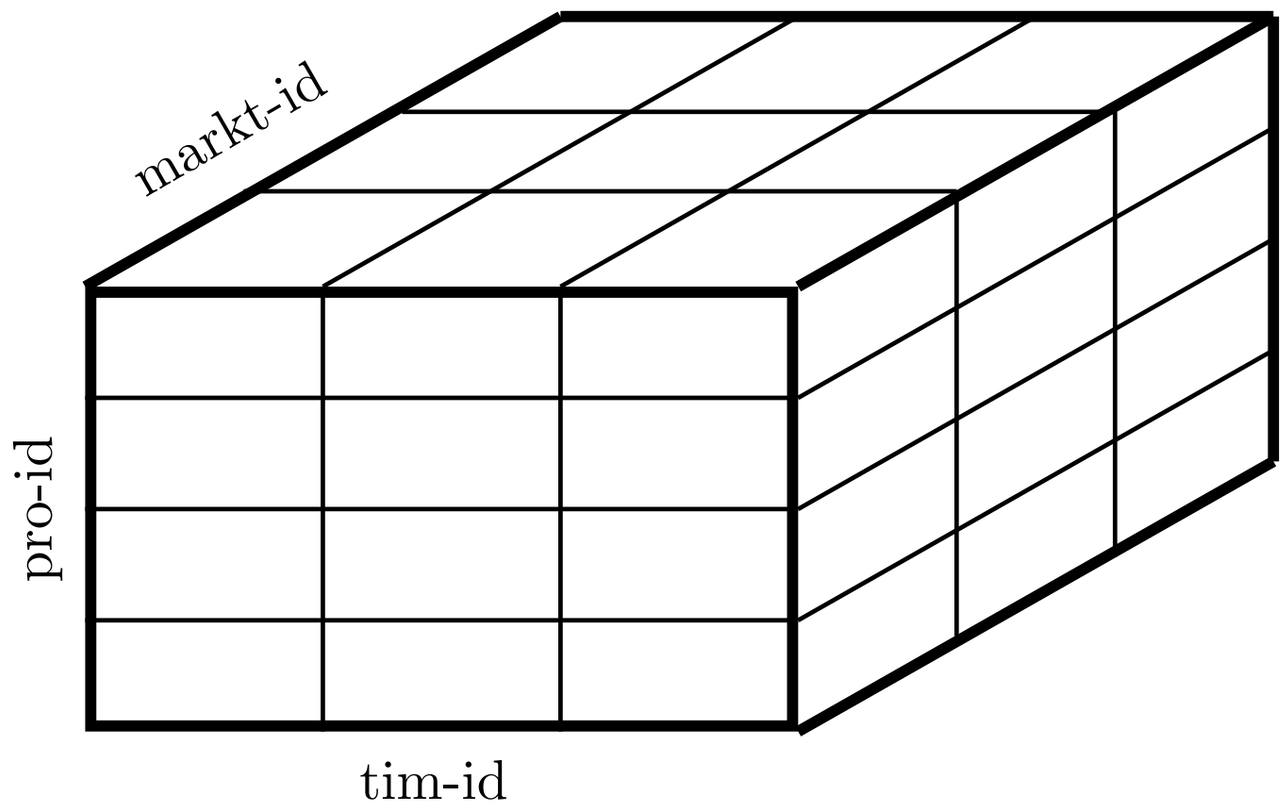
Tralie



Slicing & dicing

Dicing: group by





Slicing: where



Verkoop van produkten in kwartaal 1

Slicing & dicing

Verkoop van produkten in kwartaal 1

```
select pro_id, sum(hoev) as totqty
from Verkoop V, Periode P
where V.tim_id = P.tim_id and P.kwartaal = 1
group by pro_id;
```

pro-id	totqty
P1	6500
P2	10800
P3	1510
P4	9133

Pivoteren

```
select pro_id, maand, sum(hoev)
from Verkoop V, Periode P
where V.tim_id = P.tim_id
group by pro_id, maand;
```

	januari	juni	december
P1	6500	6503	6506
P2	10800	10803	10806
P3	1510	1513	1516
P4	9133	9136	6137

```
select pro_id, kwartaal, sum(hoev)
from Verkoop V, Periode P
where V.tim_id = P.tim_id
group by pro_id, kwartaal;
```

	1	2	4
P1	6500	6503	6506
P2	10800	10803	10806
P3	1510	1513	1516
P4	9133	9136	6137

OLAP Tools

Implementatie

- Veel gegevens, complexe queries
- Gegevens veranderen niet vaak
- Op voorhand berekenen
  - Aggregaten
  - Extra tabellen/kolommen
  - Indexen

Laden van warehouse

- Heel veel gegevens
- Uit verschillende OLTP databanken

Na T4:

```
insert into Verkoop(markt-id, pro-id, tim-id, hoev)
select 'M1', S.pro-id, S.tim-id, S.hoeveelheid
from SALES S
where S.tim-id = 'T4'
```

- Nood aan syntactische integratie: 'p01' vs. 01
- Nood aan semantische integratie: 10g vs. 10 stuks
- Nood aan cleaning

Einde

Tot slot

- Leerstof: alles behalve hoofdstuk 1
- http://www.cs.kuleuven.be/~joost/DN/
- Examen: gesloten boek
  - ± 33% theorie, mondeling
  - ± 66% oefeningen, schriftelijk