# Statistical databases in theory and practice Part II: Data models

Bo Sundgren 2010



Conceptual data model and relational data model in normalised form

## **Conceptual modelling**

- Define concepts and relations between them
- Conceptual models and data models
- Visualise models graphically

#### Rent-A-Video: first object graph



### Rent-A-Video: elaborated object graph



#### **Rent-A-Video: further aspects**



## Relations between two object types

- one-to-one, symbolised by "arrow-to-arrow"
- one-to-many, symbolised by "arrow-to-fork"
- many-to-one, symbolised by "fork-to-arrow"
- many-to-many, symbolised by "fork-to-fork"



Note: The relation is usually *not* a *flow* relation! (But you should tell what kind of relation it is.)

## **Object graphs: another example**



## Different roles of concept modelling

- Clarifying a small number of related concepts
- Information model for an application
  - defining meaning
  - basis for data design
- Corporate information model
  - for more efficient communication between people
  - basis for system integration

#### Concept model ---> Data model



Г								
Rentals	<b>FilmId</b>	CopyNr	CustomerId	RentalNr	Rental Date	Agreed ReturnDate	Returned?	Actual ReturnDate
						·'		
					_	¥		
				Custo	mers <u>Cu</u>	stomerId Na	ame Address	Discount
L	→							
FilmCopies	FilmId Co	pyNr Rented	I? NumberOfR	ents				
_								
L								
				1				
FilmTitles	s <u>FilmId</u>	Title	Category	Price	Story	Agreed ReturnDate	Returned?	Actual ReturnDate
ActorsInFilms	s FilmId	ActorNar	ne ActorsRole	InFilm				
ActorsInFilms	5 Filmld	ActorNar	ne ActorsRole	InFilm				
ActorsInFilms	s FilmId	ActorNar	ActorsRole	InFilm				
ActorsInFilms	FilmId	ActorNar	ActorsRole	InFilm				

#### Conceptual model ---> Star/cube model



#### Star model for Data Warehouse



## Multidimensional model (cube model)

during t discount ca	he year t by co tegory and fil	in copies ustomer m category_
number of rentals	number of rentals	number of rentals
number of rentals	number of rentals	number of rentals
number of rentals	number of rentals	number of rentals
number of rentals	number of rentals	number of rentals
number of rentals	number of rentals	number of rentals
	Number of during the discount call         Inumber of rentals         Inumber of rentals	Number of rentals of figduring the year t by crdiscount category and filnumber of rentalsnumber of 

CUSTOMER DIS-COUNT CATEGORY

## Part II: Extra material

Bo Sundgren 2010

## **Concept modelling: Exercises**

- B2B
  - The customers of companies are companies
  - Companies have employees (persons)
- B2C
  - The customers of companies are consumers (persons)
  - Companies have employees (persons)
- B2B+B2C
  - The customers of companies are companies or consumers (persons)
  - Companies have employees (persons)
- Hint: There are two basic object types, COMPANY and PERSON in all three examples

## Illustrate the following three examples by:

- basic object graphs (micro-level)
- relational data models (micro-level)
- multidimensional cubes (hypercubes)
- star models
- relational data models (macro-level)

Statistics	Number of foreign citizens living in Sweden and their average yearly incomes by citizenship, region, sex, and age. Years 1996-2003.
Population	foreign citizens living in Sweden at the end of year y
Population-defining properties	<ol> <li>country of residence = Sweden</li> <li>citizenship = non-Swedish</li> </ol>
Counted object	person
Related objects	
Subpopulations	crossclassification of population objects
Classification variables and value sets	<ol> <li>citizenship (country code)</li> <li>region (county.municipality)</li> <li>sex (male, female)</li> <li>age (five year age groups)</li> </ol>
Parameters	<ol> <li>number (of objects in the population)</li> <li>average (income)</li> </ol>
Reference times	year (1996,, 2003)

Example 1: Analysis of the contents and structure of a statistical message

Statistics	Number of working persons in Sweden, 16 years of age and older, living in the region (night population) by region of dwelling, region of work, sex, occupation, socio-economic status, income class, activity class of working place. Year 1990.
Population	working persons, 16+ years old, living in Sweden at the time of the reference time, t
Population-defining properties	<ol> <li>country of residence = Sweden</li> <li>working status = working</li> <li>age &gt; 15 years</li> </ol>
Counted object	person
Related objects	dwelling (where persons lives) establishments (where person work)
Subpopulations	crossclassification of population objects
Classification variables and value sets	<ol> <li>region of dwelling (country.municipality)</li> <li>region of work (county.municipality)</li> <li>sex (male, female)</li> <li>occupation (ISCO)</li> <li>socio-economic status (SEI82)</li> <li>activity class of working place (NACE)</li> </ol>
Parameters	number (of objects in the population)
Reference times	time of the population census 1990

Example 2: Analysis of the contents and structure of a statistical message

Statistics	Number migrations in Sweden by sex, age, from_region and to_region. Years 1998-2003.
Population	migration events concerning persons living in Sweden (before and/or after the event) that have taken place during the reference year, y
Population-defining properties	1. country of residence of person = Sweden
Counted object	migration event
Related objects	person who migrates, dwelling from which the person migrates, dwelling to which the person migrates
Subpopulations	crossclassification of all migration events by sex of the migrating person, region of the dwelling from which the person migrates, and region
Classification variables and value sets	<ol> <li>sex of the migrating person</li> <li>age of the migrating person</li> <li>region of dwelling from which the person migrates</li> <li>region of dwelling to which the person migrates</li> </ol>
Parameters	number (of objects in the population/subpopulation)
Reference times	years 1998-2003

#### Example 3: Analysis of the contents and structure of a statistical message