4. CSDP Step 2

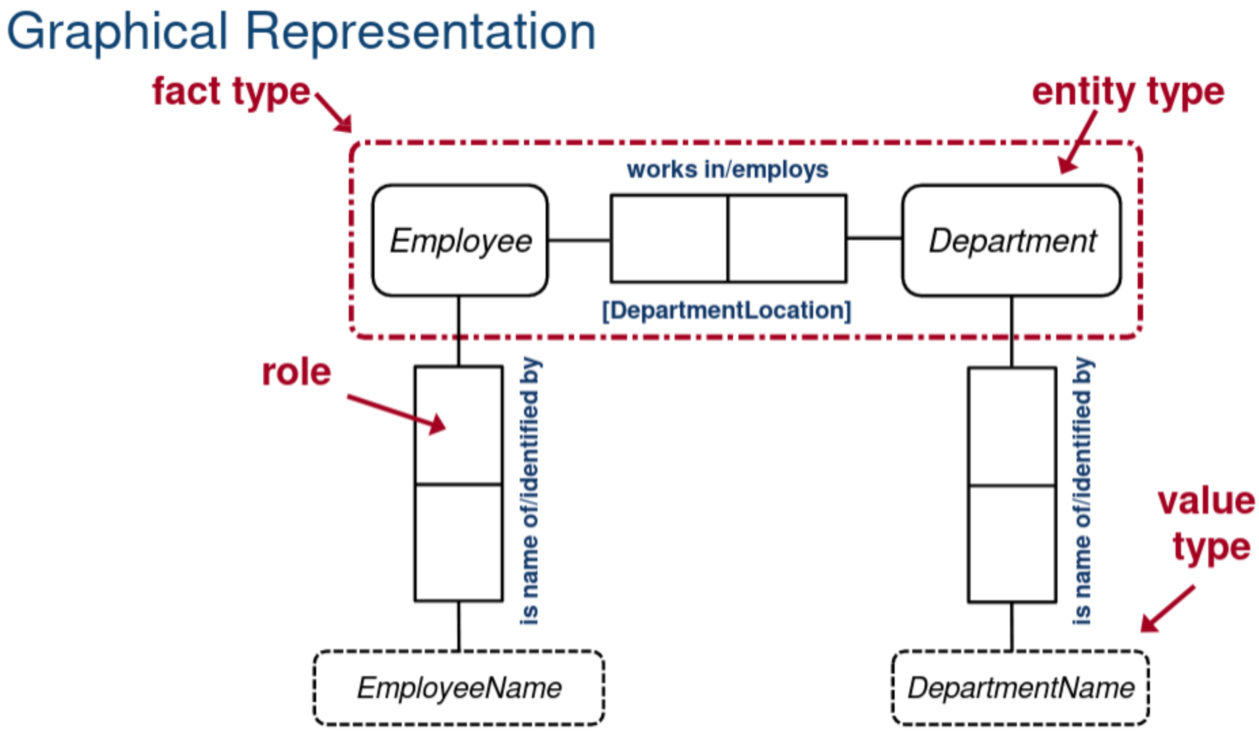
> Step 2: Draw the fact types and apply a population check.

Graphically Representing Fact Types

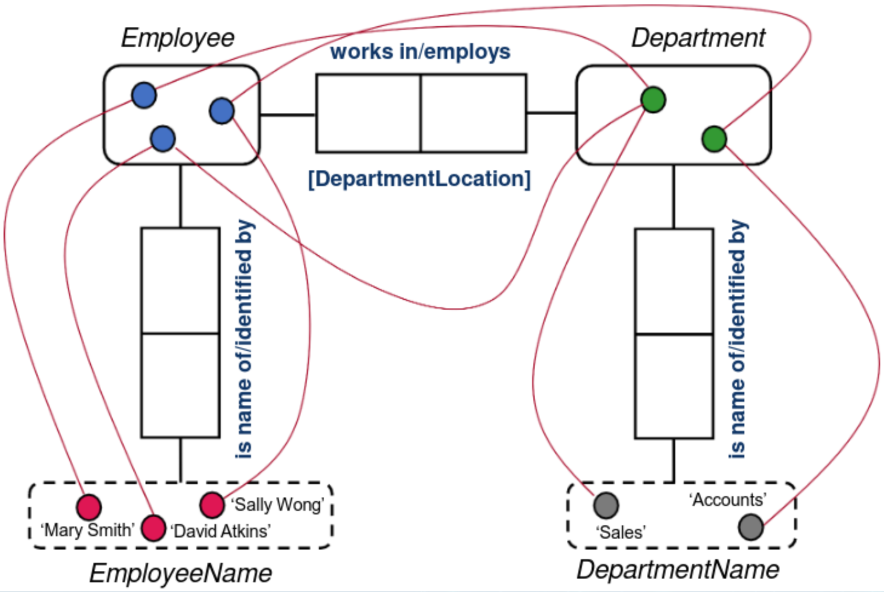
In step two, we graphically represent the fact types developed in step one. Here's an example of our two facts, which are fact type instances. At the end of step two, we should have a graphical representation of our fact types.

In the diagram outlined in red, you can see our fact type with the department location role defined. At each end of the fact type, we have our entities, which are actually entity types. We also have roles that go down to value types. It's important to note that value types have a dashed border, while entity types have a solid border.

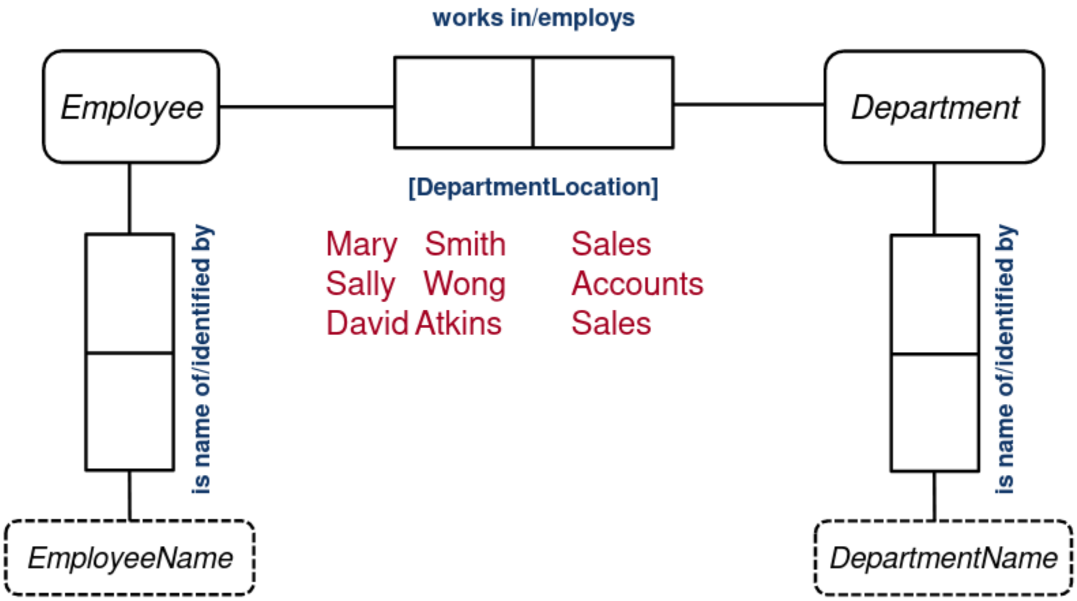
Consider the examples we've been working with, such as employees with names like Mary Smith, David Atkinson, and See Wong, and department names like sales and accounting. The dots and lines in the diagram show how they map within the diagram, and we can also show the data as text.



An Instance of the Information Base



An Instance of the Information Base



Mary Smith	Sales
Sally Wong	Accounts
David Atkins	Sales

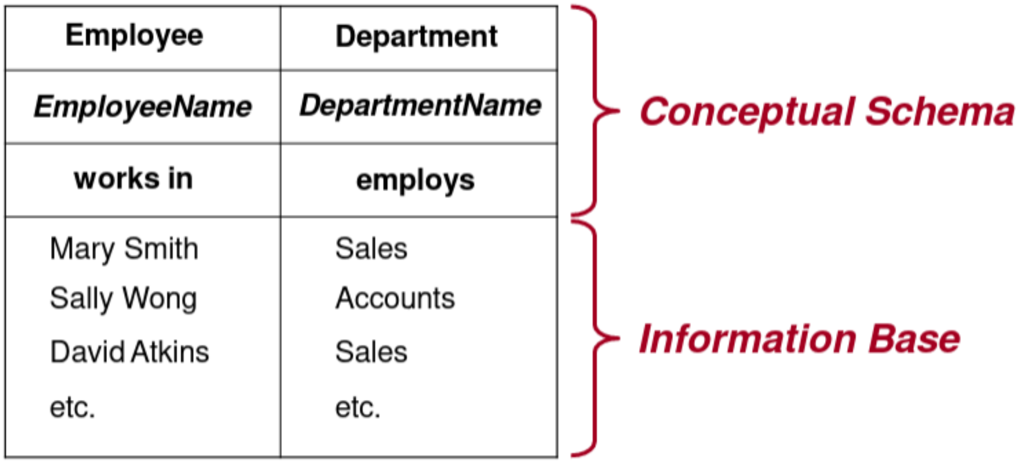
Elements of the Conceptual Schema

The conceptual schema consists of elements such as:

- Employee
- Employee name with roles
- Works in department
- Department name with role employees

We can also have example instances that make up the information base.

Tabular Representation



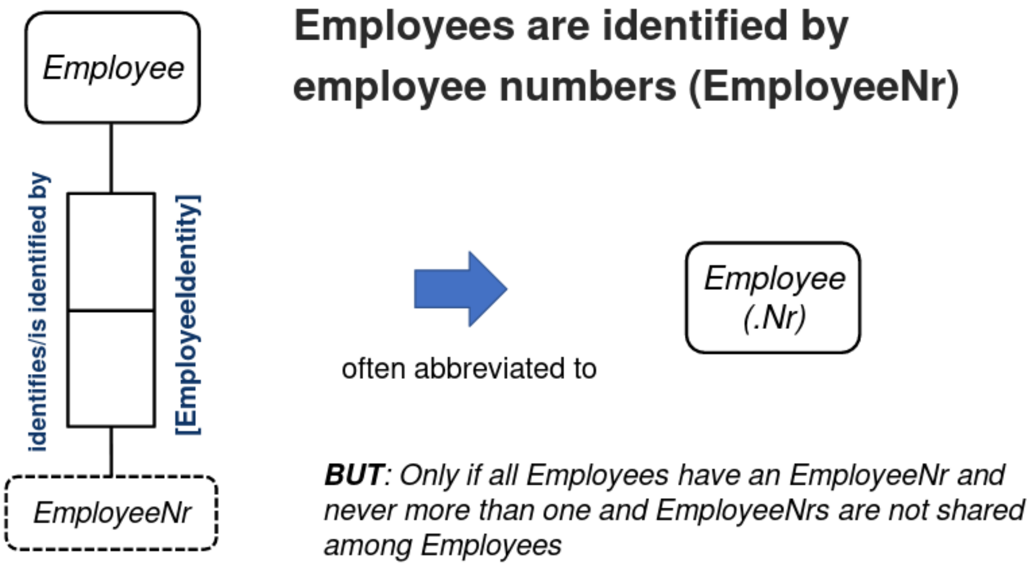
Employee	Department
EmployeeName	***DepartmentName***
works in	**employs**
Mary Smith Sally Wong David Atkins etc.	Sales Accounts Sales etc.

Conceptual Schema
Information Base

Simplifying the Diagram

We can simplify the diagram by representing an employee with their employee ID as a reference mode. If all employees have an employee number, we can represent it as an entity where the reference mode is in brackets with a preceding dot.

Graphical Representation

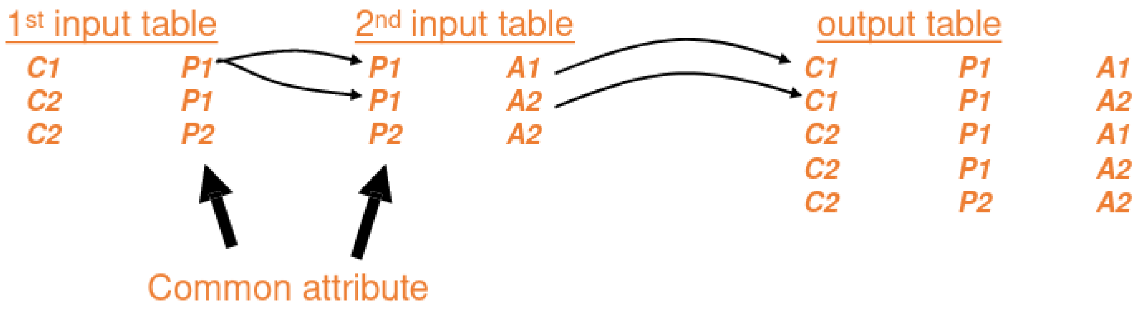


Employees are identified by employee numbers (EmployeeNr)

***BUT**: Only if all Employees have an EmployeeNr and never more than one and EmployeeNrs are not shared among Employees*

Joining Tables

In the example at the bottom, we have two tables on the left and the resulting table on the right. For each row in the first input table on the left-hand side where there is a match, we must produce a row in the output table. In the first line, C1P1 results in matching two rows in the second input table, so we end up with two rows in the output table.



1st input table

C1	P1
C2	P1
C2	P2

2nd input table

P1	A1
P1	A2
P2	A2

output table

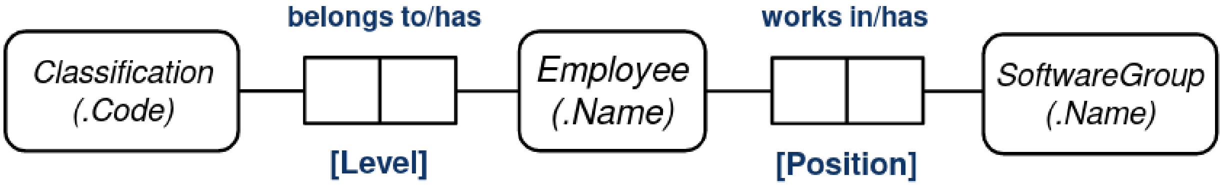
C1	P1	A1
C1	P1	A2
C2	P1	A1
C2	P1	A2
C2	P2	A2

Common attribute

Applying Step Two to Elementary Facts

Let's apply step two of our CSDP to the brief elementary facts from the previous slide. We come up with a diagram where each entity has a reference mode:

- Classification with reference mode code
- Employee with reference mode name
- Software group with reference mode name

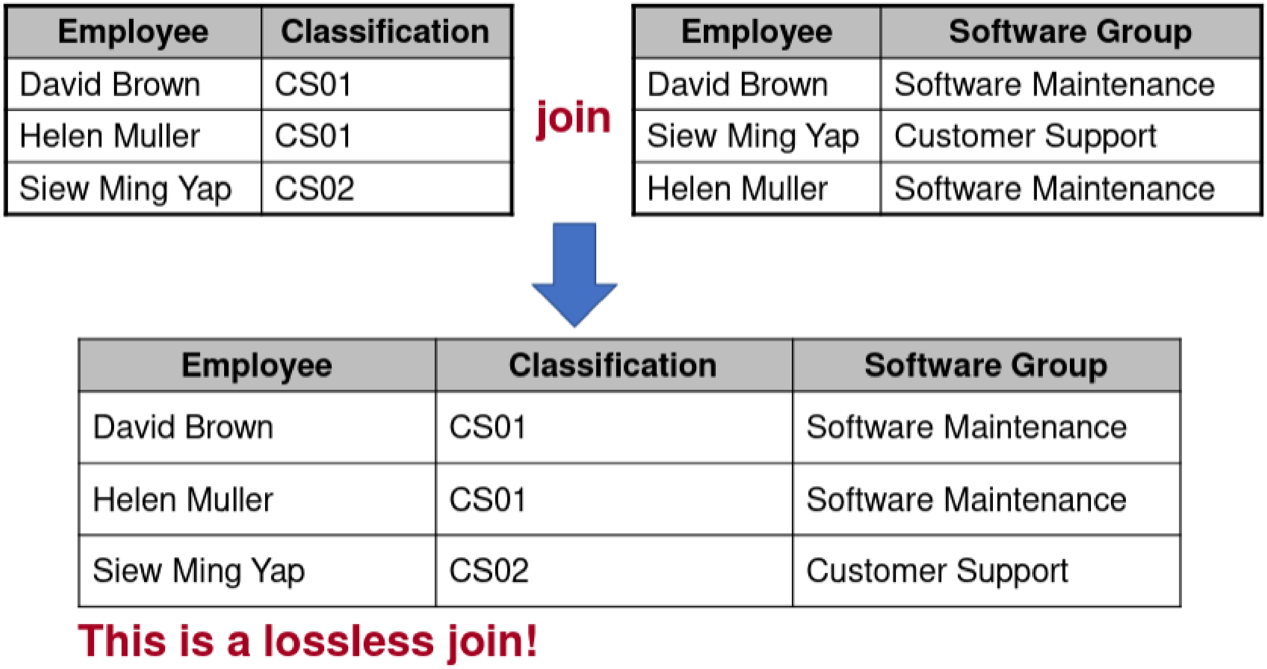


CS01	David Brown	David Brown	Software Maintenance
CS01	Helen Muller	Helen Muller	Software Maintenance
CS02	Siew Ming Yap	Siew Ming Yap	Customer Support

We also have the roles: level and position.

If we apply the two separate pieces in the information base (the employee classification table and the employee software group table) and join them, we can see that we don't lose any data.

This is called a lossless join.



Employee	Classification
David Brown	CS01
Helen Muller	CS01
Siew Ming Yap	CS02

join

Employee	Software Group
David Brown	Software Maintenance
Siew Ming Yap	Customer Support
Helen Muller	Software Maintenance

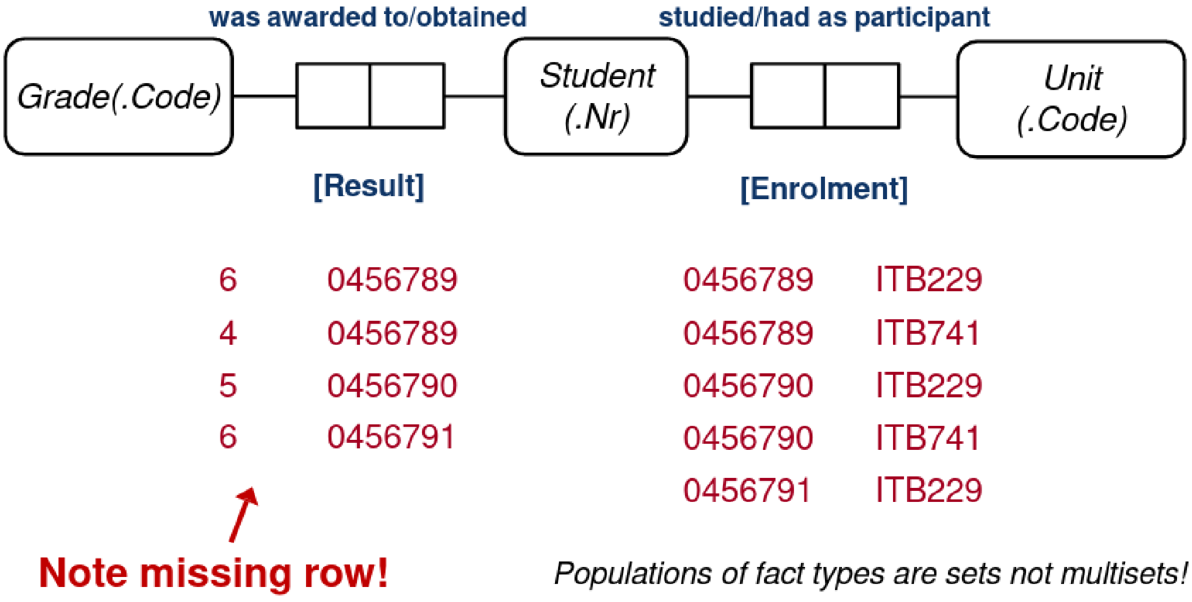
Employee	Classification	Software Group
David Brown	CS01	Software Maintenance
Helen Muller	CS01	Software Maintenance
Siew Ming Yap	CS02	Customer Support

This is a lossless join!

Identifying Ternary Relationships

In another example, we have three entities with reference modes:

- Grade with reference mode code
- Student with reference mode number
- Unit with reference mode code



6	0456789
4	0456789
5	0456790
6	0456791

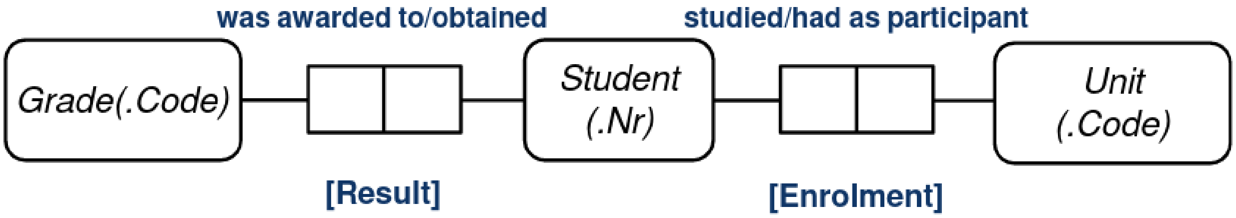
0456789	ITB229
0456789	ITB741
0456790	ITB229
0456790	ITB741
0456791	ITB229

Note missing row!

Populations of fact types are sets not multisets!

We have two roles called result and one called enrollment.

When we map our information base, we see that we have one row missing on the left-hand side, equivalent to the role result. If we apply the process of joining the two tables at the top of the slide, we end up with two additional incorrect facts compared to the original information base. This tells us that the binary relationship is not working, and it is actually a ternary relationship.



was awarded to/obtained

studied/had as participant

Grade(.Code)

Student (.Nr)

Unit (.Code)

[Result]

[Enrolment]

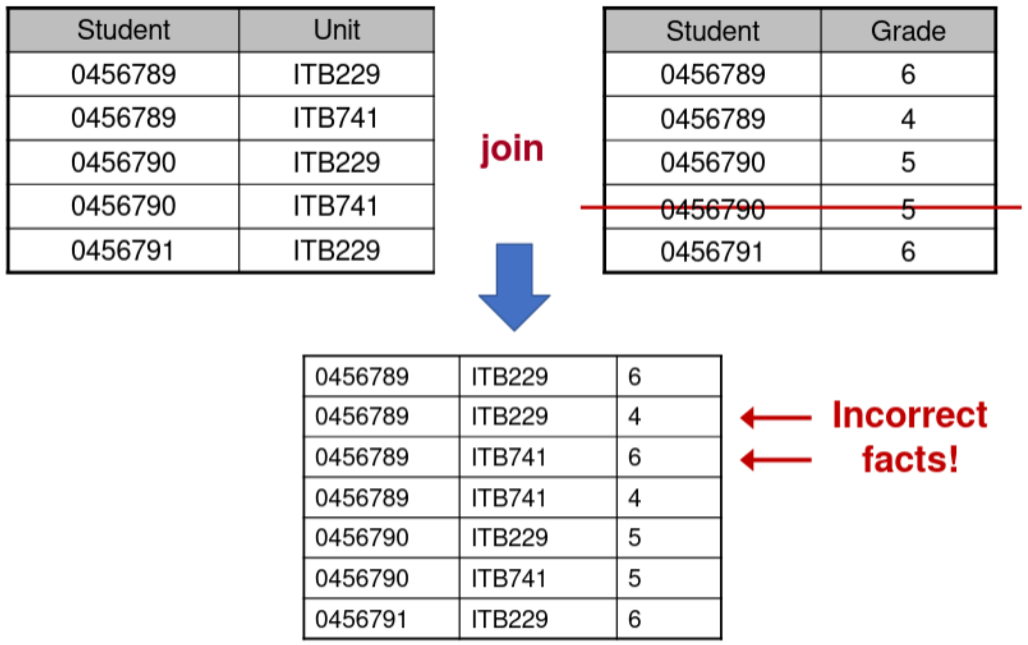
6	0456789
4	0456789
5	0456790
6	0456791

0456789	ITB229
0456789	ITB741
0456790	ITB229
0456790	ITB741
0456791	ITB229

Note missing row!

Populations of fact types are sets not multisets!

Join



Student	Unit
0456789	ITB229
0456789	ITB741
0456790	ITB229
0456790	ITB741
0456791	ITB229

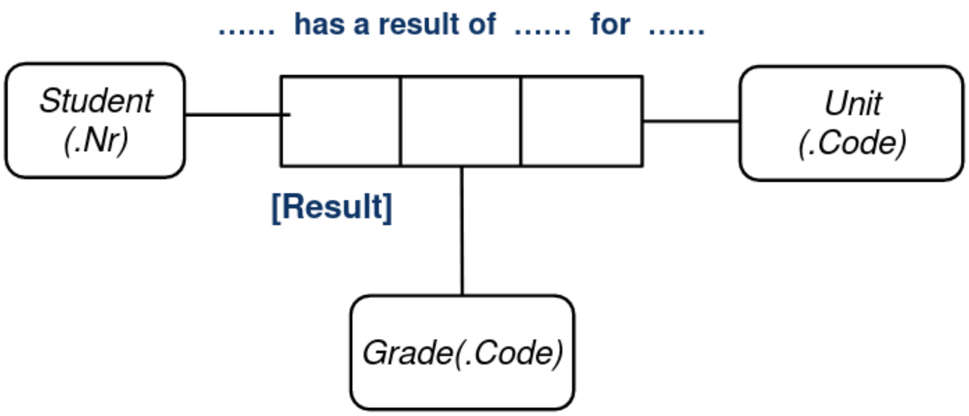
join

Student	Grade
0456789	6
0456789	4
0456790	5
~~0456790~~	~~5~~
0456791	6

0456789	ITB229	6
0456789	ITB229	4
0456789	ITB741	6
0456789	ITB741	4
0456790	ITB229	5
0456790	ITB741	5
0456791	ITB229	6

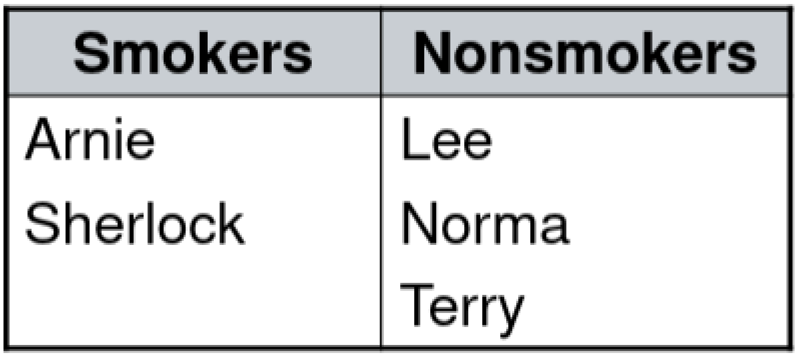
Incorrect facts!

Correct Conceptual Schema



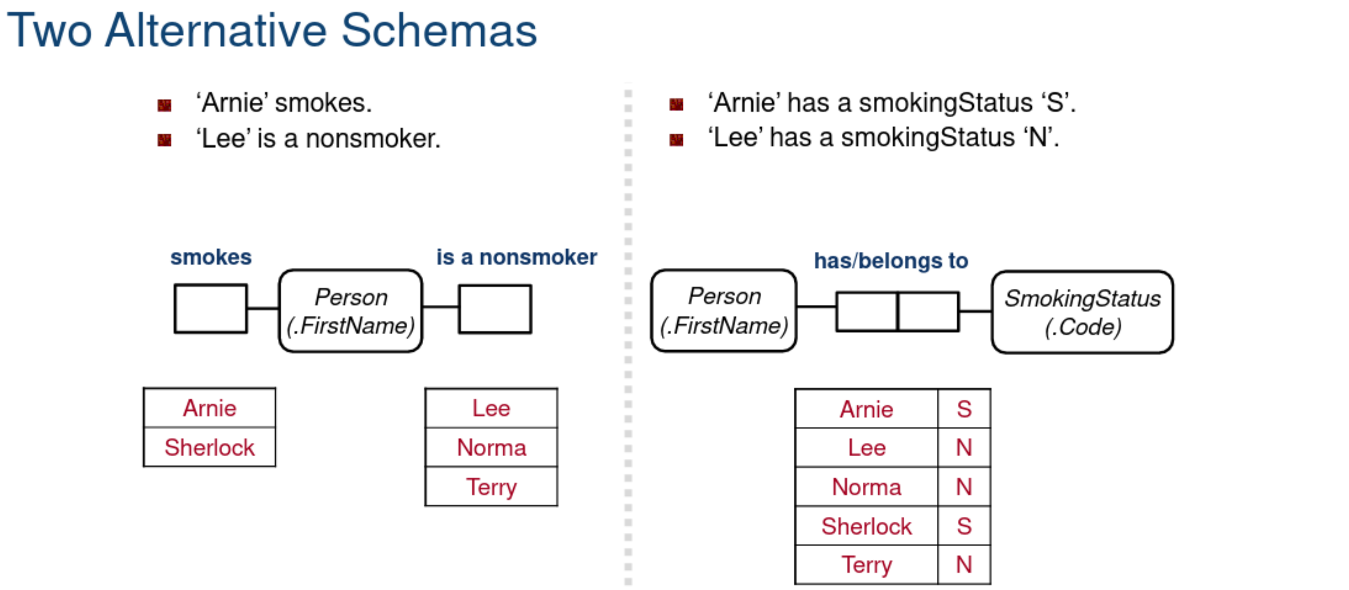
Roles and Entity Types

In the last example, we compare two alternative schemas.



Smokers	Nonsmokers
Arnie	Lee
Sherlock	Norma
	Terry

The one on the left has one entity type and two roles, while the one on the right has two entity types and one role. In the schema on the right, we've encoded the smoking status.



Two Alternative Schemas

- 'Arnie' smokes.
- 'Lee' is a nonsmoker.

Arnie
Sherlock

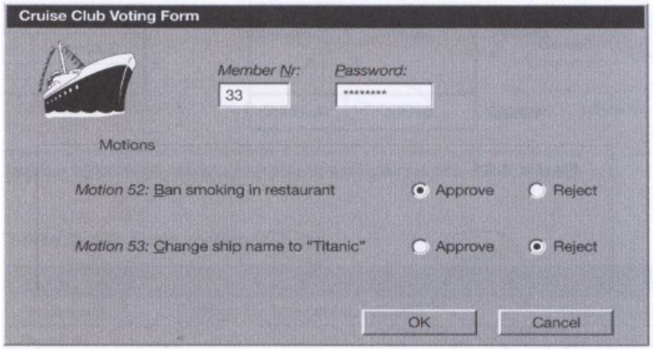
Lee
Norma
Terry

- 'Arnie' has a smokingStatus 'S'.
- 'Lee' has a smokingStatus 'N'.

Arnie	S
Lee	N
Norma	N
Sherlock	S
Terry	N

Halpin, T. and Morgan, T. "Information Modeling and Relational Databases", 2nd edition, Morgan Kaufmann, 2008. **Page 86.**

Voting System Example



Cruise Club Voting Form

Member Nr: 33

Password: ********

Motions

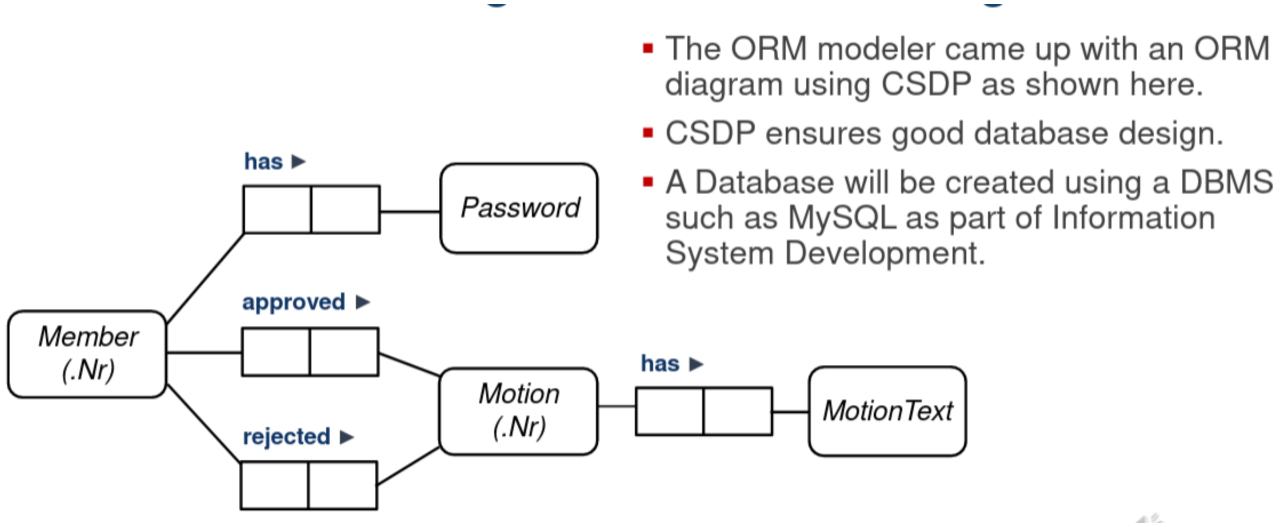
Motion 52: Ban smoking in restaurant ● Approve ○ Reject

Motion 53: Change ship name to "Titanic" ○ Approve ● Reject

This form is used by a club for voting on motions the members are supposed to vote.

Using the information presented in previous images, we've identified the elementary facts and fact types for a voting system. The ORM model shows our fact types and entity types:

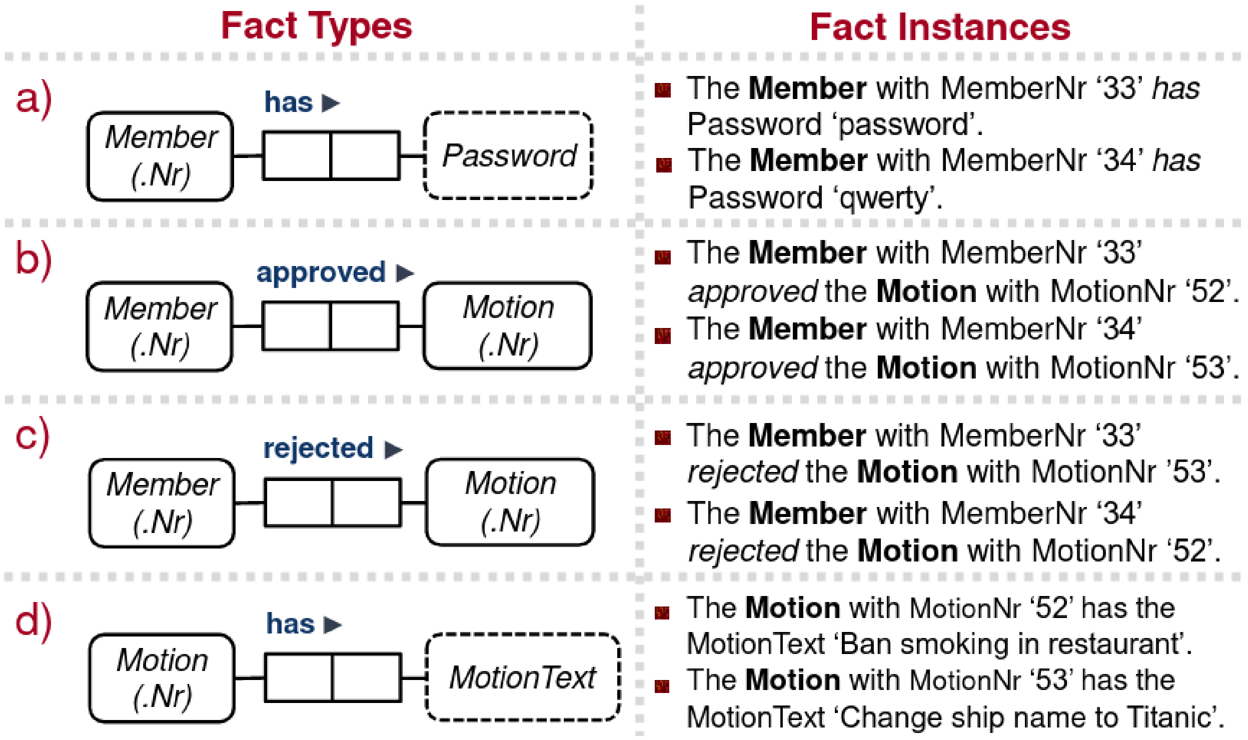
- Entity types: member, motion
- Value types: password, motion text
- Reference modes: member number, motion number
- Fact types: approved, rejected



- The ORM modeler came up with an ORM diagram using CSDP as shown here.
- CSDP ensures good database design.
- A Database will be created using a DBMS such as MySQL as part of Information System Development.

An alternate solution for the voting system uses a ternary fact type between member and motion, replacing approved and rejected. In this ternary fact type, we can have non-splittable facts, which can be interpreted as follows:

- The member with member number 33 voted on the motion with motion number 52
- The member with member number 33 voted to approve



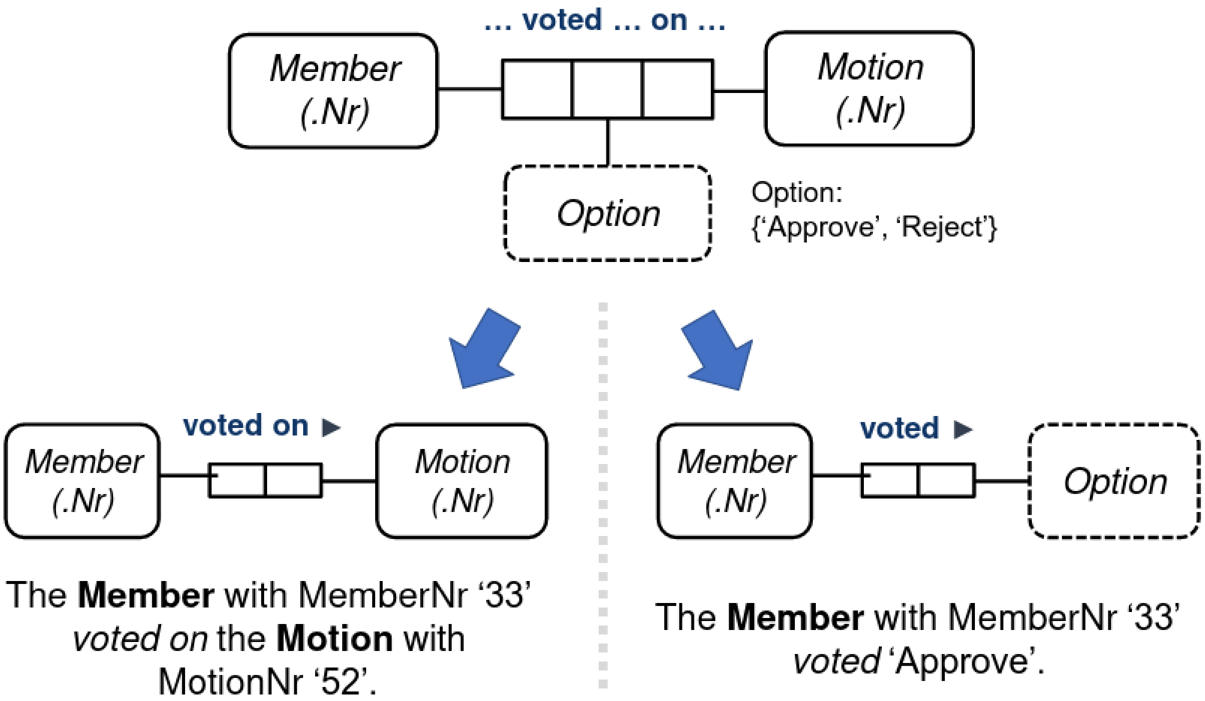
Fact Types | **Fact Instances**

a)
- The **Member** with MemberNr ‘33’ *has* Password ‘password’.
- The **Member** with MemberNr ‘34’ *has* Password ‘qwerty’.

b)
- The **Member** with MemberNr ‘33’ *approved* the **Motion** with MotionNr ‘52’.
- The **Member** with MemberNr ‘34’ *approved* the **Motion** with MotionNr ‘53’.

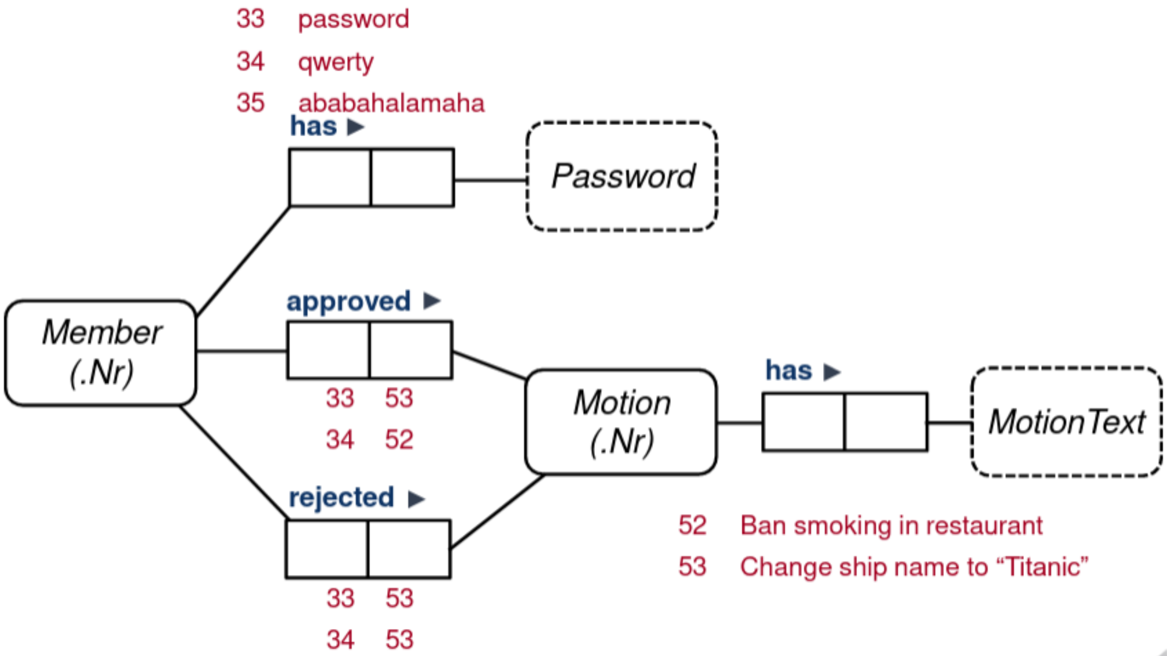
c)
- The **Member** with MemberNr ‘33’ *rejected* the **Motion** with MotionNr ’53’.
- The **Member** with MemberNr ‘34’ *rejected* the **Motion** with MotionNr ‘52’.

d)
- The **Motion** with MotionNr ‘52’ has the MotionText ‘Ban smoking in restaurant’.
- The **Motion** with MotionNr ‘53’ has the MotionText ‘Change ship name to Titanic’.



The **Member** with MemberNr '33' *voted on* the **Motion** with MotionNr '52'.

The **Member** with MemberNr '33' *voted* 'Approve'.

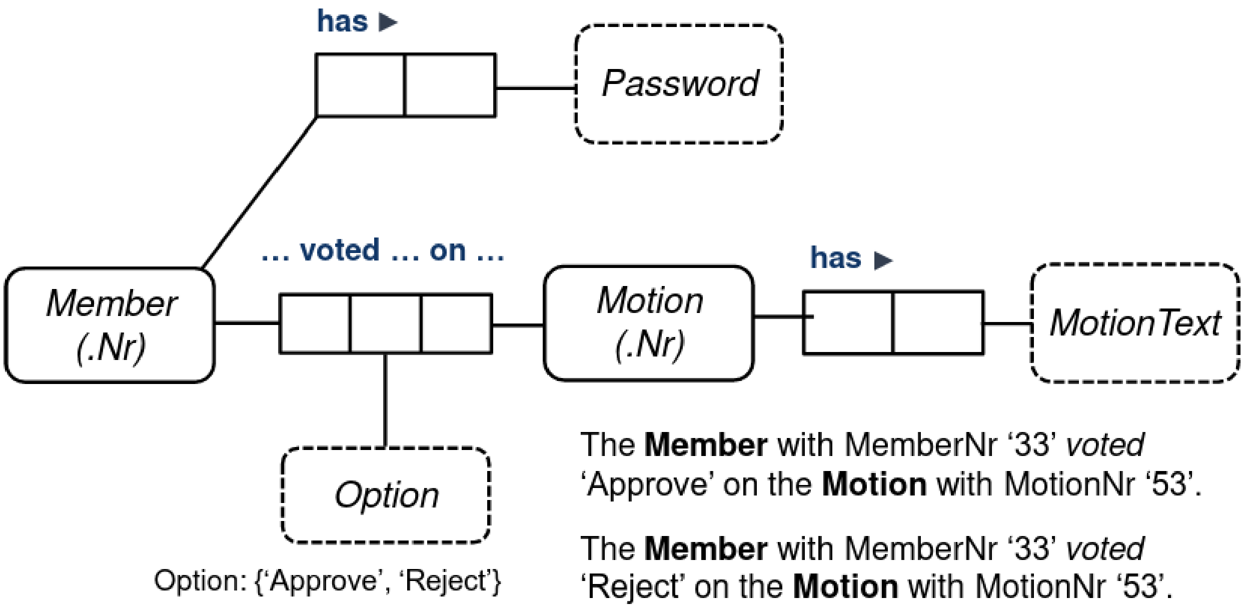


33	password
34	qwerty
35	ababahalamaha

33	53
34	52

33	53
34	53

52	Ban smoking in restaurant
53	Change ship name to "Titanic"



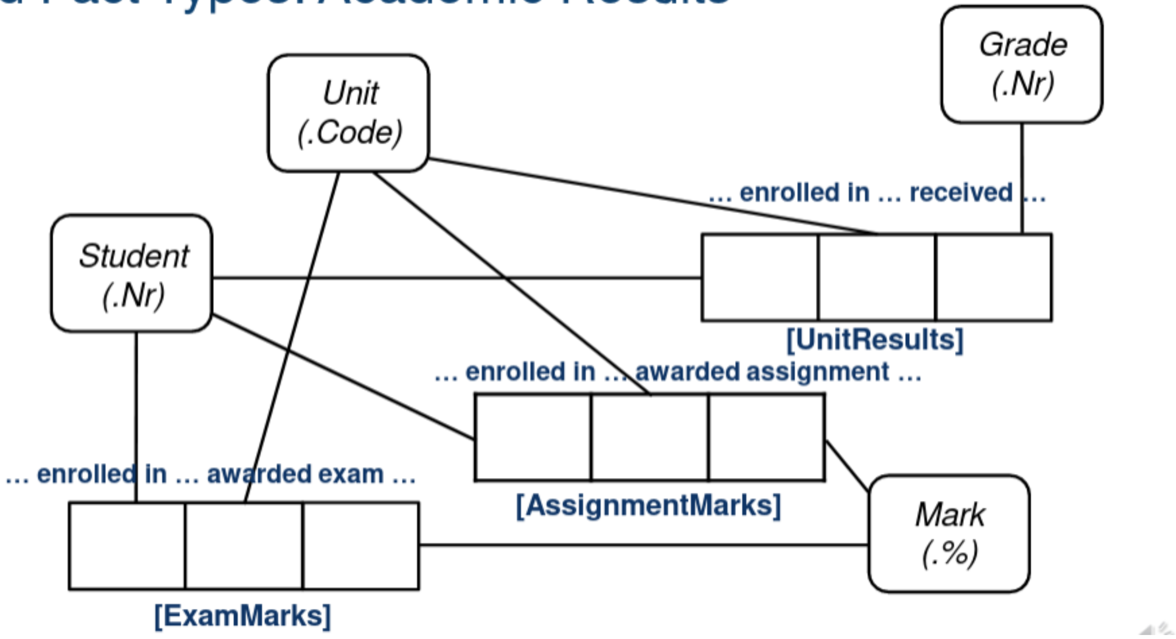
The **Member** with MemberNr '33' *voted* 'Approve' on the **Motion** with MotionNr '53'.

The **Member** with MemberNr '33' *voted* 'Reject' on the **Motion** with MotionNr '53'.

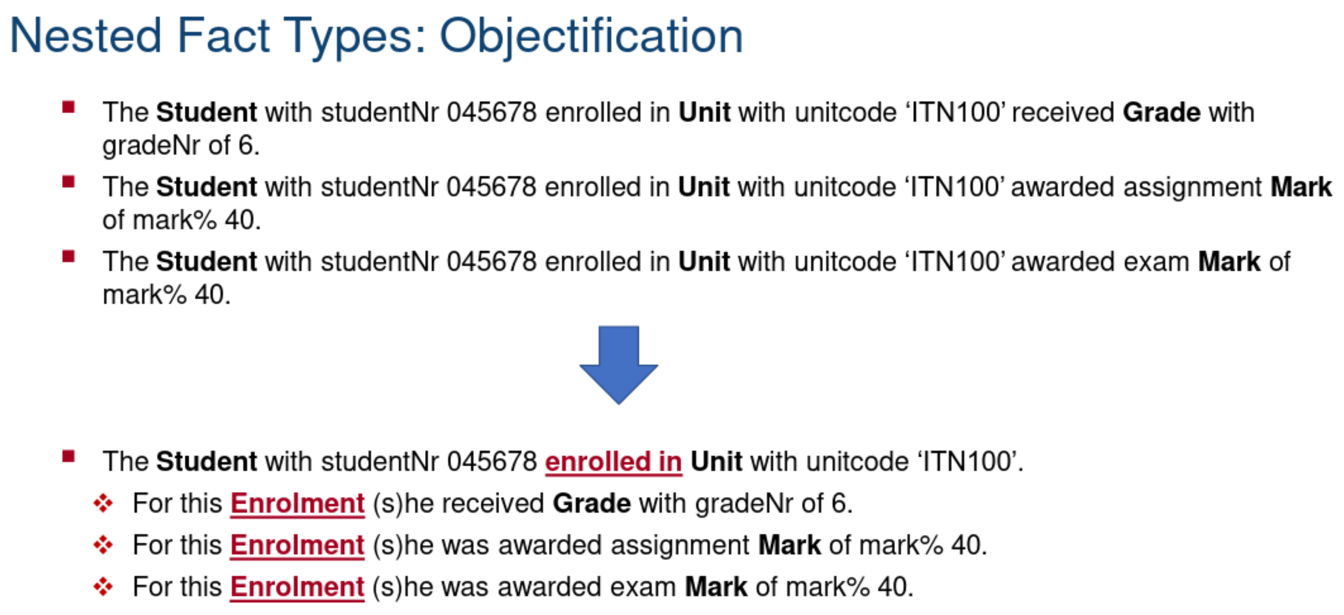
Nested Fact Types

In this example, we look at nested fact types based on academic results. The resulting diagram shows our three ternary roles and four entity types, each with a reference mode.

Nested Fact Types: Academic Results


Grade (.Nr)
Unit (.Code)
… enrolled in … received …
Student (.Nr)
[UnitResults]
… enrolled in … awarded assignment …
… enrolled in … awarded exam …
[AssignmentMarks]
Mark (.%)
[ExamMarks]

Within our elementary facts, the term "enrolled in unit with unit code" is common in all three facts. We can objectify this and relabel it as "enrollment" so that the information about graded assignment mark and exam mark is for a particular enrollment.

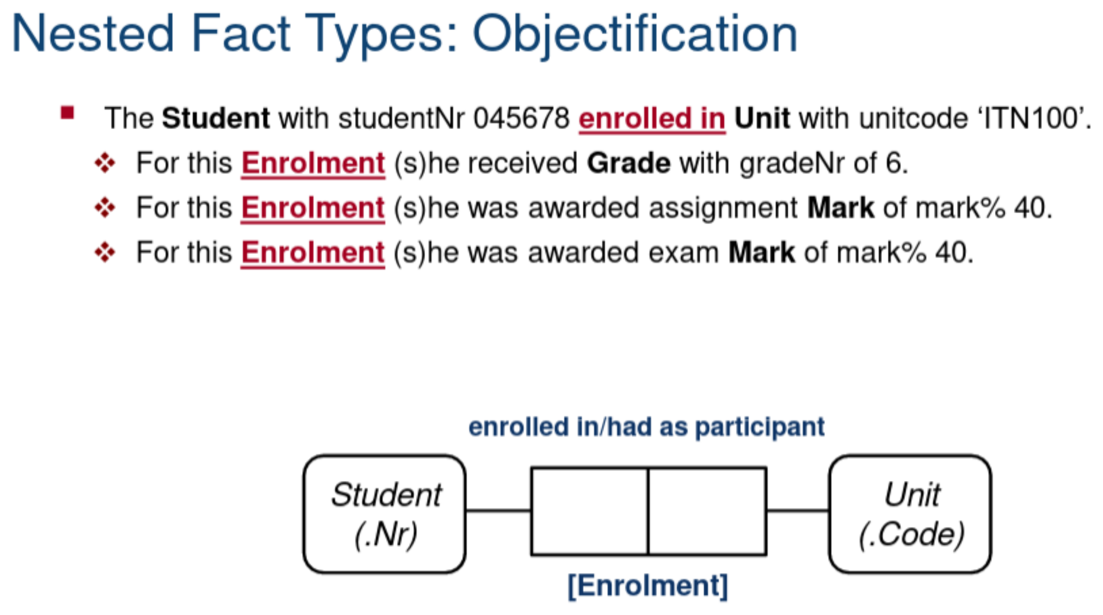


Nested Fact Types: Objectification

- The **Student** with studentNr 045678 enrolled in **Unit** with unitcode 'ITN100' received **Grade** with gradeNr of 6.
- The **Student** with studentNr 045678 enrolled in **Unit** with unitcode 'ITN100' awarded assignment **Mark** of mark% 40.
- The **Student** with studentNr 045678 enrolled in **Unit** with unitcode 'ITN100' awarded exam **Mark** of mark% 40.

- The **Student** with studentNr 045678 **enrolled in** **Unit** with unitcode 'ITN100'.
  - ❖ For this **Enrolment** (s)he received **Grade** with gradeNr of 6.
  - ❖ For this **Enrolment** (s)he was awarded assignment **Mark** of mark% 40.
  - ❖ For this **Enrolment** (s)he was awarded exam **Mark** of mark% 40.

The nested fact type becomes the objectification and is graphically represented between the student and unit entity types. The other roles and entity types are connected to the nested fact type.

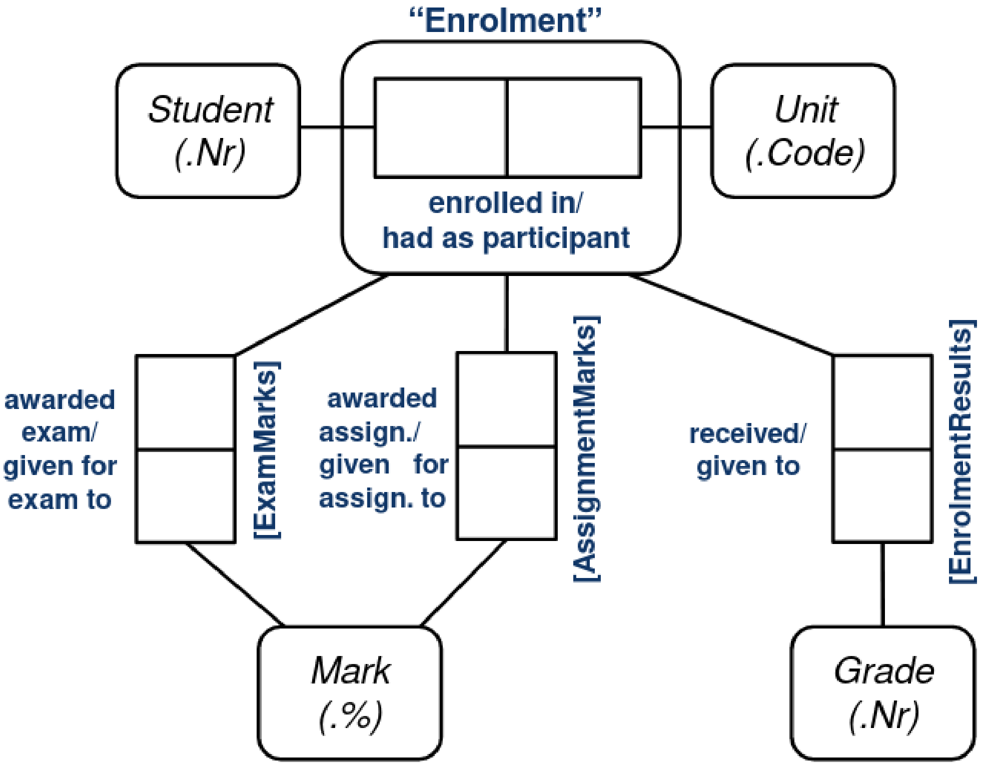


Nested Fact Types: Objectification

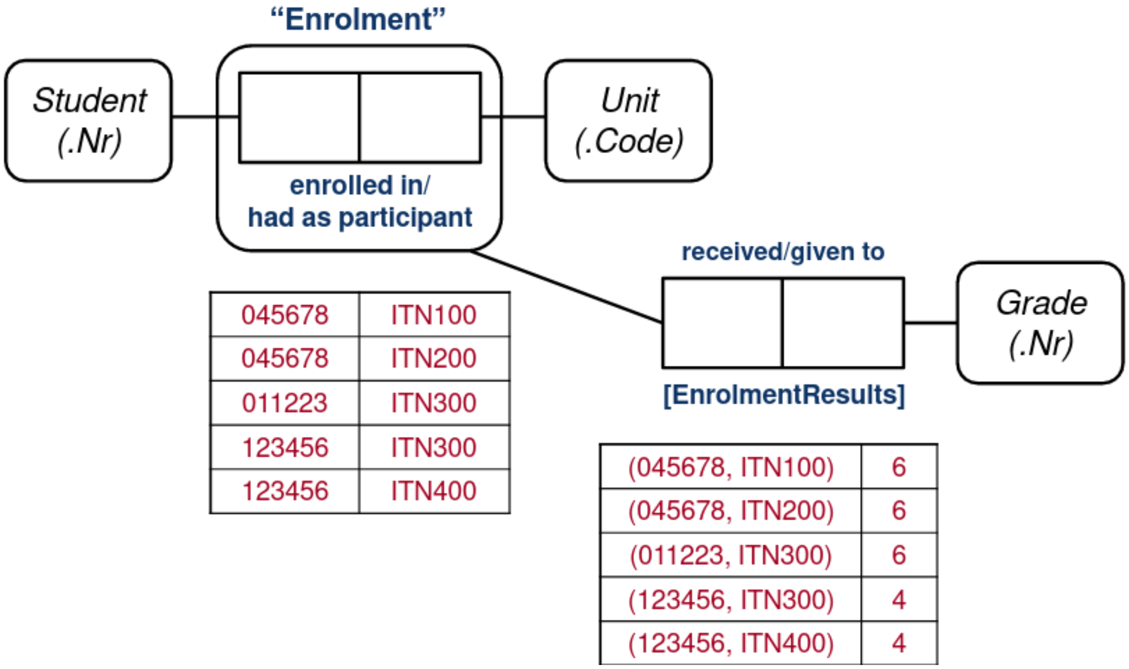
- The **Student** with studentNr 045678 **enrolled in** **Unit** with unitcode 'ITN100'.
  - ❖ For this **Enrolment** (s)he received **Grade** with gradeNr of 6.
  - ❖ For this **Enrolment** (s)he was awarded assignment **Mark** of mark% 40.
  - ❖ For this **Enrolment** (s)he was awarded exam **Mark** of mark% 40.

To check the validity of the nested fact type, we can populate it with data and confirm that it works as expected.

Nested Fact Types: Academic Results



Populate the (Nested) Fact Types



045678	ITN100
045678	ITN200
011223	ITN300
123456	ITN300
123456	ITN400

(045678, ITN100)	6
(045678, ITN200)	6
(011223, ITN300)	6
(123456, ITN300)	4
(123456, ITN400)	4

See Also

5. CSDP Step 3