



Functional Dependencies – Part 1

Introduction



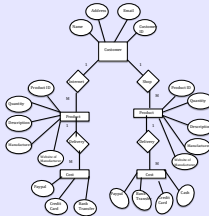
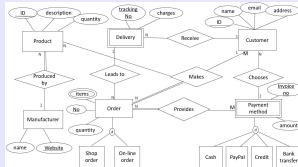
Database Design Quality

- A fundamental question in database design:

What constitutes a “well-designed” database schema?

- We have learnt that:
 - A database design often starts with building an EER model.
 - An EER model can then be translated to a relational database schema.
- However, such an EER model may not be “*perfect*”. Instead, it is common to have many different EER models for the same application.

Database Design Quality - Examples ¹



¹ Previous COMP2400/6240 students' solutions for an EER modelling question



Database Design Quality

- Some desirable properties of a “well-designed” database schema
 - **Completeness**
Has all relevant information been captured?
 - **Redundancy freeness**
Has the doubling of relevant information been avoided (if possible)?
 - **Consistent understanding**
Is the meaning of all relevant information consistent?
Is the meaning of NULL clear?
 - Does not apply
 - Unknown
 - Known but absent
 - **Performance**
Can the database schema lead to the good performance for given tasks?

Motivating Example

- Suppose that we want to store the enrolment information (i.e., *course no*, *semester* and *unit*) of students (i.e., *name*, *student id* and *date of birth*) in a relational database.
- Is the design of the relation ENROLMENT good?**

ENROLMENT					
Name	<u>StudentID</u>	DoB	<u>CourseNo</u>	<u>Semester</u>	Unit
Tom	123456	25/01/1988	COMP2400	2010 S2	6
Tom	123456	25/01/1989	COMP8740	2011 S2	12
Michael	123458	21/04/1985	COMP2400	2009 S2	6
Michael	123458	21/04/1985	COMP8740	2011 S2	12
Fran	123456	11/09/1987	COMP2400	2009 S2	8

Motivating Example – Data Inconsistency

- Any inconsistency problems with these tuples?



Tom	123456	25/01/1988	COMP2400	2010 S2	6
Tom	123456	25/01/1989	COMP8740	2011 S2	12

The same student has different DoBs. *This seems unreasonable.*



Michael	123458	21/04/1985	COMP2400	2009 S2	6
Fran	123456	11/09/1987	COMP2400	2009 S2	8

There are different units for the same course in the same semester. *That should not happen.*



Tom	123456	25/01/1989	COMP8740	2011 S2	12
Fran	123456	11/09/1987	COMP2400	2009 S2	8

The different students have the same ID. *This is unacceptable.*



Motivating Example – Data Redundancy

- Any redundancy problems with these tuples?



Michael	123458	21/04/1985	COMP2400	2009 S2	6
Michael	123458	21/04/1985	COMP8740	2011 S2	12

There exists redundant information about students.



Tom	123456	25/01/1989	COMP8740	2011 S2	12
Michael	123458	21/04/1985	COMP8740	2011 S2	12

There exists redundant information about courses.

Motivating Example – Update Anomalies

- What could happen to update operations (e.g., insert, delete and update)?

ENROLMENT					
Name	<u>StudentID</u>	DoB	<u>CourseNo</u>	<u>Semester</u>	Unit
Tom	123456	25/01/1988	COMP2400	2010 S2	6
Tom	123456	25/01/1988	COMP8740	2011 S2	12
Michael	123458	21/04/1985	COMP2400	2009 S2	6
Michael	123458	21/04/1985	COMP8740	2011 S2	12
Fran	123456	11/09/1987	COMP2400	2009 S2	6

- Modification anomalies:** If changing the DoB of Michael, then ...
- Insertion anomalies:** If inserting a new course COMP3000, then ...
- Deletion anomalies:** If deleting the enrolled course COMP2400 of Fran, then ...



Database Design Issues

- We have seen the following database design issues so far:
 - Data inconsistency
 - Data redundancy
 - Update anomalies

ENROLMENT					
Name	<u>StudentID</u>	DoB	<u>CourseNo</u>	<u>Semester</u>	Unit
Tom	123456	25/01/1988	COMP2400	2010 S2	6
Tom	123456	25/01/1989	COMP8740	2011 S2	12
Michael	123458	21/04/1985	COMP2400	2009 S2	6
Michael	123458	21/04/1985	COMP8740	2011 S2	12
Fran	123456	11/09/1987	COMP2400	2009 S2	8

- **Can we avoid these issues when designing a database?**



Database Design Issues - Motivating Example

- We may fix those database design issues through breaking a relation into smaller relations.

ENROLMENT					
Name	<u>StudentID</u>	DoB	<u>CourseNo</u>	<u>Semester</u>	Unit
Tom	123456	25/01/1988	COMP2400	2010 S2	6
Tom	123456	25/01/1988	COMP8740	2011 S2	12
Michael	123458	21/04/1985	COMP2400	2009 S2	6
Michael	123458	21/04/1985	COMP8740	2011 S2	12
Fran	123457	11/09/1987	COMP2400	2009 S2	6

- For example, each tuple in ENROLMENT represents **three** different facts:
 - 1 Information about students
 - 2 Information about courses
 - 3 Course enrolment of students



Database Design Issues - Motivating Example

ENROLMENT					
Name	<u>StudentID</u>	DoB	<u>CourseNo</u>	<u>Semester</u>	Unit
Tom	123456	25/01/1988	COMP2400	2010 S2	6
Tom	123456	25/01/1988	COMP8740	2011 S2	12
Michael	123458	21/04/1985	COMP2400	2009 S2	6
Michael	123458	21/04/1985	COMP8740	2011 S2	12
Fran	123457	11/09/1987	COMP2400	2009 S2	6



STUDENT		
Name	<u>StudentID</u>	DoB
Tom	123456	25/01/1988
Michael	123458	21/04/1985
Fran	123457	11/09/1987

COURSE	
<u>CourseNo</u>	Unit
COMP2400	6
COMP8740	12

ENROL		
<u>StudentID</u>	<u>CourseNo</u>	<u>Semester</u>
123456	COMP2400	2010 S2
123456	COMP8740	2011 S2
123458	COMP2400	2009 S2
123458	COMP8740	2011 S2
123457	COMP2400	2009 S2