Problems of Redundancy

Gary KL Tam

Department of Computer Science
Swansea University

Redundancy

See the following spreadsheet example from Hoffer's book.
Primary key is underlined.

empid	name	dept	salary	course	date_completed
100	Margaret Simpson	Marketing	42000	SPSS	19/6/96
100	Margaret Simpson	Marketing	42000	Surveys	20/7/96
140	Alan Beeton	Accounting	39000	Tax Acc	21/8/96
110	Chris Lucero	IT	41500	SPSS	22/9/96
110	Chris Lucero	IT	41500	C++	23/10/96
190	Lorenzo Davis	Finance	38000	Investments	24/11/96
150	Susan Martin	Marketing	38500	SPSS	25/11/96
150	Susan Martin	Marketing	38500	TQM	12/1/97

Think

Why is it a bad relation?

Problem 1

empid	name	dept	salary	course	date_completed
100	Margaret Simpson	Marketing	42000	SPSS	19/6/96
100	Margaret Simpson	Marketing	42000	Surveys	20/7/96
140	Alan Beeton	Accounting	39000	Tax Acc	21/8/96
110	Chris Lucero	IT	41500	SPSS	22/9/96
110	Chris Lucero	IT	41500	C++	23/10/96
190	Lorenzo Davis	Finance	38000	Investments	24/11/96
150	Susan Martin	Marketing	38500	SPSS	25/11/96
150	Susan Martin	Marketing	38500	TQM	12/1/97

Insert Problem

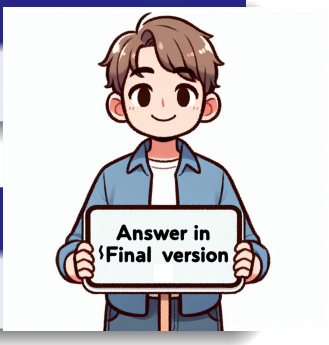
?

Delete Problem

?

Update Problem

?



Problem 2

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

Insert Problem

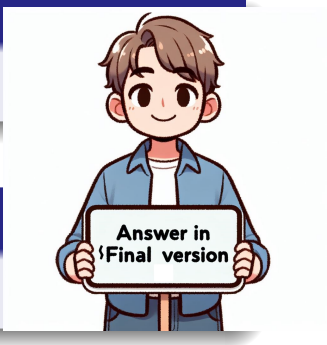
?

Delete Problem

?

Update Problem

?



Problem 2 - solution

How can we resolve these?

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

Split tables...

course	staff
CS307	Phil Grant
CS328	Phil Grant
CS217	Mark Jones
CS121	Mark Jones
CS250	Gary Tam

staff	room
Phil Grant	312
Mark Jones	209
Gary Tam	206

Update Problem

?

Problem 2 - solution

staff	course	room
Phil Grant	CS307	312
Phil Grant	CS328	312
Mark Jones	CS217	209
Mark Jones	CS121	209
Gary Tam	CS250	206

course	staff_id
CS307	1
CS328	1
CS217	2
CS121	2
CS250	3

staff_id	staff	room
1	Phil Grant	312
2	Mark Jones	209
3	Gary Tam	206

These may sound extreme!

But consider a big database and the user has no ideas how many duplicates are there.

Problem 1 - solution

empid	name	dept	salary	course	date_completed
100	Margaret Simpson	Marketing	42000	SPSS	19/6/96
100	Margaret Simpson	Marketing	42000	Surveys	20/7/96
140	Alan Beeton	Accounting	39000	Tax Acc	21/8/96
110	Chris Lucero	IT	41500	SPSS	22/9/96
110	Chris Lucero	IT	41500	C++	23/10/96
190	Lorenzo Davis	Finance	38000	Investments	24/11/96
150	Susan Martin	Marketing	38500	SPSS	25/11/96
150	Susan Martin	Marketing	38500	TQM	12/1/97

Splitting tables...

empid	name	dept	salary
100	Margaret Simpson	Marketing	42000
140	Alan Beeton	Accounting	39000
110	Chris Lucero	IT	41500
190	Lorenzo Davis	Finance	38000
150	Susan Martin	Marketing	38500

empid	course	date_completed
100	SPSS	19/6/96
100	Surveys	20/7/96
140	Tax Acc	21/8/96
110	SPSS	22/9/96
110	C++	23/10/96
190	Investments	24/11/96
150	SPSS	25/11/96
150	TQM	12/1/97

- Redundancy creates insert, delete, update problems.
- Split tables can help - the basis of normalization process.
- The previous process is very informal.
- We will study how to do this formally.