

A decorative graphic on the left side of the slide, consisting of a network of white lines and circles on a blue gradient background, resembling a circuit board or a tree structure.

WEEK 9

NORMALIZATION – FINAL NOTES

STUDENT OBJECTIVES

- Upon completion of this video, you should be able to:
 - Use the Lossless Join Property to check if you have normalized (broken up) your tables correctly
 - Identify situations where you might want to denormalize your tables.

LOSSLESS JOIN PROPERTY

- Guarantees that no spurious tuples are generated
- A decomposition $D = \{R1, R2\}$ of R has the **lossless join property** with respect to a set of functional dependencies F on R if and only if either:
 - $R1 \cap R2$ ($R1$ intersect $R2$) $\rightarrow R1$, that is: all attributes common to both $R1$ and $R2$ functionally determine ALL the attributes in $R1$ **OR**
 - $R1 \cap R2$ ($R1$ intersect $R2$) $\rightarrow R2$, that is: all attributes common to both $R1$ and $R2$ functionally determine ALL the attributes in $R2$

i.e. the shared key is functionally dependent on $r1$ or $r2$.

QUESTION: Draw the functional dependencies for the following table: BIG_TABLE

R: BIG TABLE:

SSN	PNUMBER	HOURS	ENAME	PNAME
-----	---------	-------	-------	-------

Suppose now we break it down into the following tables again:

R1: EMP HOURS:

ENAME	HOURS
-------	-------

R2: EMP PROJ1:

SSN	PNUMBER	PNAME	ENAME
-----	---------	-------	-------

QUESTION: Is this a lossless decomposition, why or why not? ✗

intercept. $R1 \cap R2 = \{ENAME\}$
if correctly split the table, then ENAME would either
functionally determine R1 or R2.

- Helps us avoid doing “dumb” things like splitting this:

SSN	PNUMBER	HOURS	ENAME	PNAME
1	A	5	Smith	Alpha
1	B	4	Smith	Beta
2	C	10	Jones	Cappa
3	A	12	Cook	Alpha
3	B	33	Cook	Beta
4	B	23	Aziz	Beta
4	C	45	Aziz	Cappa
4	D	23	Aziz	Delta

into this:

ENAME	HOURS
Smith	5
Smith	4
Jones	10
Cook	12
Cook	33
Aziz	23
Aziz	45
Aziz	23

SSN	PNUMBER	ENAME	PNAME
1	A	Smith	Alpha
1	B	Smith	Beta
2	C	Jones	Cappa
3	A	Cook	Alpha
3	B	Cook	Beta
4	B	Aziz	Beta
4	C	Aziz	Cappa
4	D	Aziz	Delta

$R1 = \{ENAME, HOURS\}$ and $R2 = \{SSN, PNUMBER, ENAME, PNAME\}$
 But ENAME does NOT \rightarrow R2 and ENAME does NOT \rightarrow R1

Now look at the next decomposition:

R1:

PNUMBER	PNAME
---------	-------

R2:

SSN	PNUMBER	HOURS	ENAME
-----	---------	-------	-------

QUESTION: Is this a lossless decomposition, why or why not? ✓

PNUMBER → R1.

DEPENDENCY PRESERVATION PROPERTY

- ensures that each functional dependency is represented in some individual SINGLE relation after the big relation is decomposed into smaller relations.
- **Note:** Relations in third normal form will preserve the dependency but not necessarily in BCNF (Boyce-Codd Normal Form)
- **NOTE:** Boyce-Codd is another level of normalization, as is fourth, fifth and sixth level normalization but they are not very common.

DENORMALIZATION

- While a normalized database is the “Gold Standard”, you will have to do JOINS to answer most queries and for large amounts of data, JOINS are expensive (take a long time to run).
- Thus you might want to **denormalize** your tables to make your queries more efficient
- Great video that explains why you might do **Denormalization (from: Database Denormalization – Programming Foundations: Databases with Simon Allardice:)**
<https://www.lynda.com/MyPlaylist/Watch/17844556/438440?autoplay=true>

REVIEW

- These videos are great review of normalization:
 - <https://www.lynda.com/MyPlaylist/Watch/17844556/78157?autoplay=true>
(Database Normalization – SQL Server 2008 Essential Training with Simon Allardice)
 - <http://www.youtube.com/watch?v=ygfikznRjpw>

Student

StudentNum FIV LV CID G CN

T_1

SIV FV LV G

T_2

CID CN G