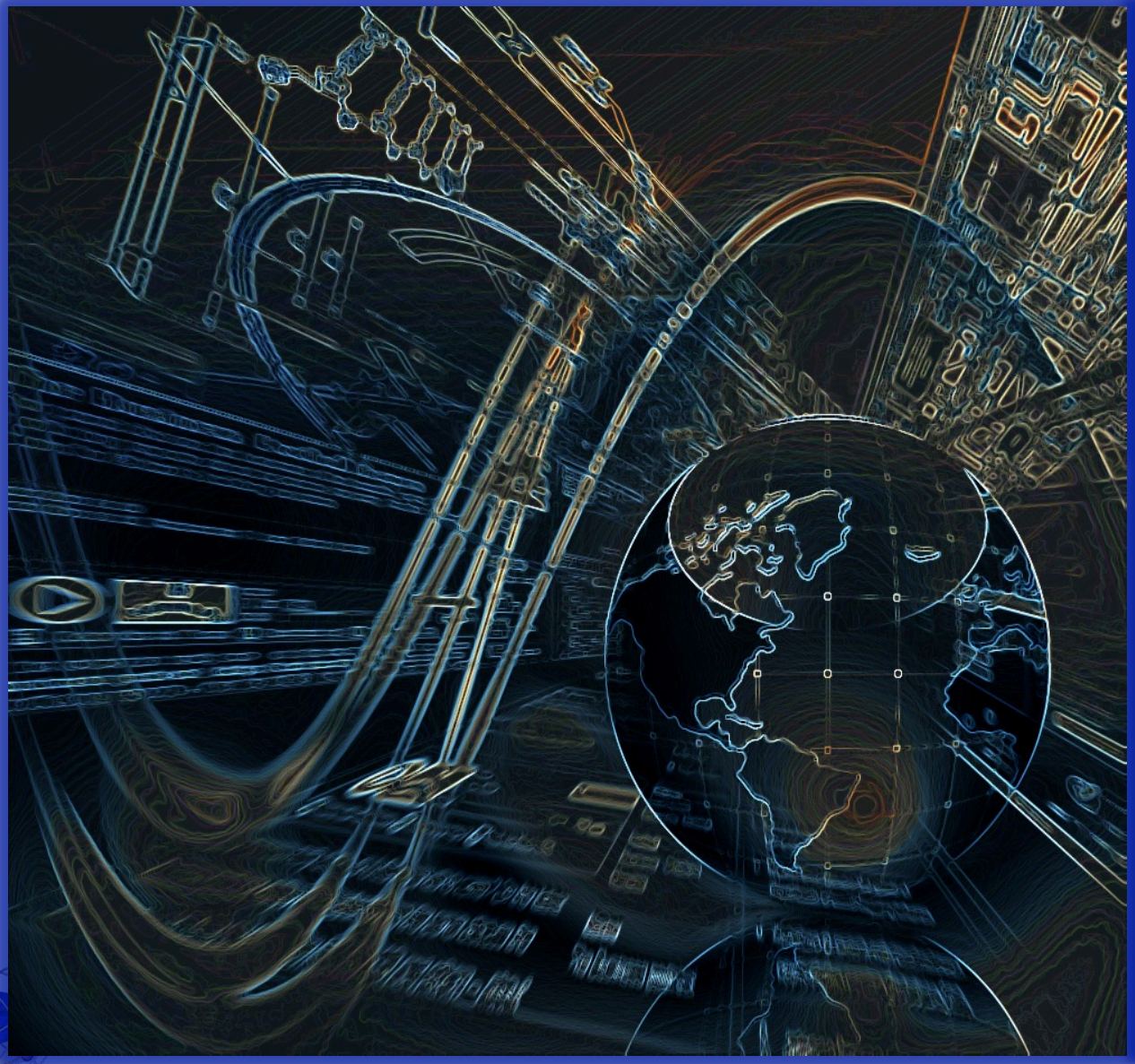
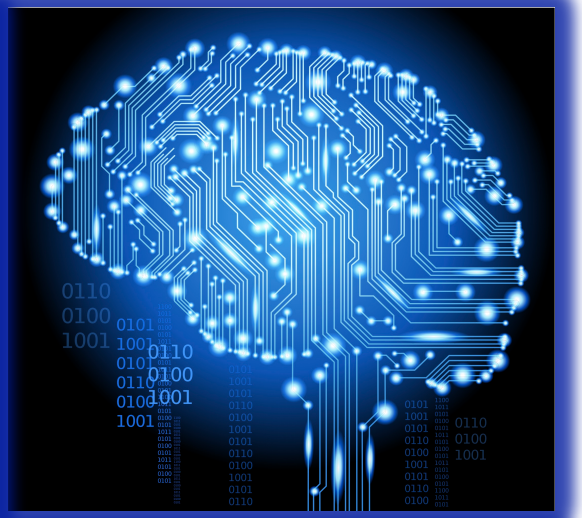
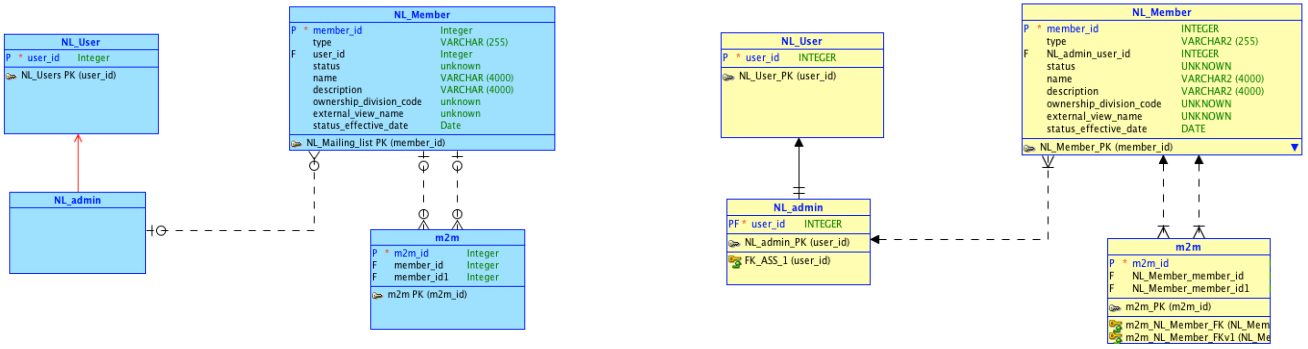
Relational Normalization Theory

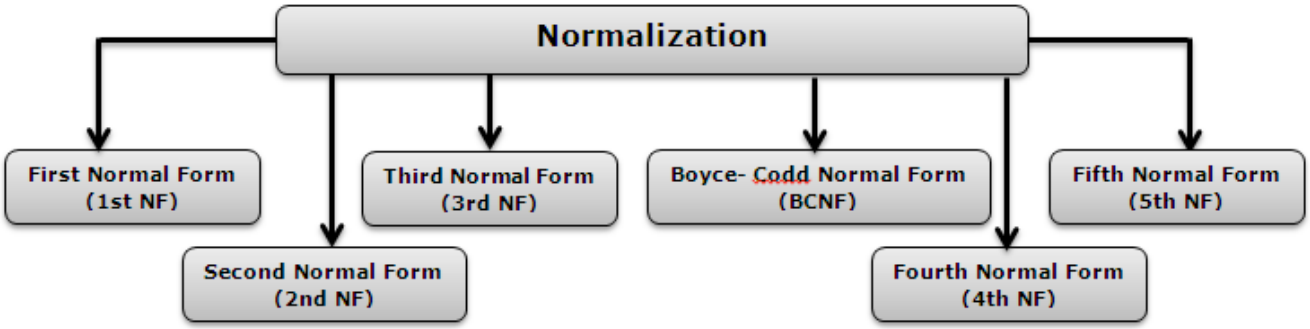




Designing Databases Using Normalization

On No!!! There's another way to design a good set of tables besides Conceptual and Logical Modeling.





Normalization

FDs, Keys, 1NF, 2NF, 3NF, BCNF, MVDs, F+, A+, FD Preservation, Lossless Join …

See also http://en.wikipedia.org/wiki/Database_normalization

A Badly Designed Table*

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	DNAME	LOC
7369	SMITH	CLERK	7902	17-DEC-80	824		20	RESEARCH	DALLAS
7499	ALLEN	SALESMAN	7698	20-FEB-81	1600	300	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	22-FEB-81	1250	500	30	SALES	CHICAGO
7566	JONES	MANAGER	7839	02-APR-81	2975		20	RESEARCH	DALLAS
7654	MARTIN	SALESMAN	7698	28-SEP-81	1250	1400	30	SALES	CHICAGO
7698	BLAKE	MANAGER	7839	01-MAY-81	2850		30	SALES	CHICAGO
7782	CLARK	MANAGER	7839	09-JUN-81	2450		10	ACCOUNTING	NEW YORK
7788	SCOTT	ANALYST	7566	09-DEC-82	3000		20	RESEARCH	DALLAS
7839	KING	none		17-NOV-81	5000		10	ACCOUNTING	NEW YORK
7844	TURNER	SALESMAN	7698	08-SEP-81	1500		30	SALES	CHICAGO
7876	ADAMS	CLERK	7788	12-JAN-83	1100		20	RESEARCH	DALLAS
7900	JAMES	CLERK	7698	03-DEC-81	950		30	SALES	CHICAGO
7902	FORD	ANALYST	7566	03-DEC-81	3000		20	RESEARCH	DALLAS
7934	MILLER	CLERK	7782	23-JAN-82	1300		10	ACCOUNTING	NEW YORK

This table was create with the following SQL: create table emp_dept as (select e., dname, loc from emp e, dept d where e.deptno = d.deptno)

What's wrong with this table?

- **Insertion Anomalies**

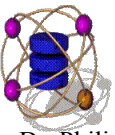
  Can't insert information about Department 40 unless it has an employee. So where do you store information about Department 40?

- **Deletion Anomalies**

  If you delete EMPNOs 7782, 7839, and 7934, you lose all information about Department 10.

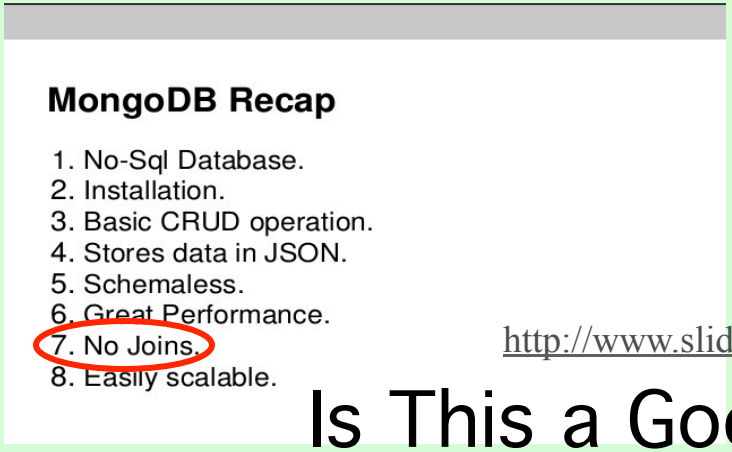
- **Update Anomalies**

  If you change CLARK's DNAME to 'SALES' and leave his DEPTNO=10, you need to make sure the DNAME for DEPT 10 is changed in every tuple.



What does “No Joins” Mean in MongoDB?


```
{"company" : [
{  "name":"ACCOUNTING","identifier":"10","location":"NEW YORK",
  "employees":[{  "name":"CLARK","job":"MANAGER","salary":2450, "manager":{ "name":"KING","salary":
5000,"hiredate":"17-11-1981"} ,"hiredate":"09-06-1981"} ,
          {  "name":"KING","job":"PRESIDENT","salary":5000,"manager":"","hiredate":"17-11-1981"} ]} ,
{  "name":"OPERATIONS","identifier":"40","location":"BOSTON",
  "employees":[]} ,
{  "name":"RESEARCH","identifier":"20","location":"DALLAS",
  "employees":[{  "name":"ADAMS","job":"CLERK","salary":1100,  "manager":{ "name":"SCOTT","salary":
3000,"hiredate":"09-12-1982"} ,"hiredate":"12-01-1983"} ,
          {  "name":"FORD","job":"ANALYST","salary":3000, "manager":{ "name":"JONES","salary":2975,"hiredate":"02-04-1981"} ,"hiredate":"03-12-1981"} ,
          {  "name":"JONES","job":"MANAGER","salary":2975,"manager":{ "name":"KING","salary":5000,"hiredate":"17-11-1981"}  ,"hiredate":"02-04-1981"} ,
          {  "name":"SCOTT","job":"ANALYST","salary":3000,"manager":{ "name":"JONES","salary":2975,"hiredate":"02-04-1981"} ,"hiredate":"09-12-1982"} ,
          {  "name":"SMITH","job":"CLERK","salary":800,   "manager":{ "name":"FORD","salary":3000,"hiredate":"03-12-1981"}  ,"hiredate":"17-12-1980"} ]} ,
{  "name":"SALES","identifier":"30","location":"CHICAGO",
  "employees":[{  "name":"ALLEN","job":"SALESMAN","salary":1600, "manager":{ "name":"BLAKE","salary":
2850,"hiredate":"01-05-1981"} ,"hiredate":"20-02-1981"} ,
          {  "name":"BLAKE","job":"MANAGER","salary":2850,  "manager":{ "name":"KING","salary":5000,"hiredate":"17-11-1981"}  ,"hiredate":"01-05-1981"} ,
          {  "name":"JAMES","job":"CLERK","salary":950,     "manager":{ "name":"BLAKE","salary":2850,"hiredate":"01-05-1981"} ,"hiredate":"03-12-1981"} ,
          {  "name":"MARTIN","job":"SALESMAN","salary":1250,"manager":{ "name":"BLAKE","salary":2850,"hiredate":"01-05-1981"} ,"hiredate":"28-09-1981"} ,
          {  "name":"TURNER","job":"SALESMAN","salary":1500,"manager":{ "name":"BLAKE","salary":2850,"hiredate":"01-05-1981"} ,"hiredate":"08-09-1981"} ,
          {  "name":"WARD","job":"SALESMAN","salary":1250,  "manager":{ "name":"BLAKE","salary":2850,"hiredate":"01-05-1981"} ,"hiredate":"22-02-1981"} ]}
]}
```

http://www.slideshare.net/amitthakkar01/get-expertise-with-mongo-db

Is This a Good Thing?

Functional Dependencies

Informal definition of a Functional Dependency

- **Require that the value for a certain set of attributes uniquely determines the value for another set of attributes.**
- **A functional dependency is a generalization of the notion of a *key*.**

Formal definition of a Functional Dependency

Let R be a relation and

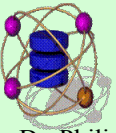
$$\alpha \subseteq R \quad and \quad \beta \subseteq R$$

The functional dependency

$$\alpha \rightarrow \beta$$

holds on R if and only if for any tuples t_1 and t_2 that agree on the attributes α, they also agree on the attributes β.

That is, $t_1[\alpha] = t_2[\alpha] \Rightarrow t_1[\beta] = t_2[\beta]$



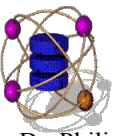
Keys

Keys

- ***Superkey*** - *K* is a superkey for relation *R* if and only if $K \rightarrow R$
- ***Candidate Key*** - *K* is a candidate key for *R* if and only if

  $K \rightarrow R$ *(i.e., K is a super key)*, and for no $\alpha \subset K$, $\alpha \rightarrow R$

  (i.e., K is minimal with respect to the number of attributes of which it is composed)
- ***Prime attributes*** - attributes that belong to any candidate key
- ***Primary Key*** – Pick one from the candidate keys



Armstrong's Axioms

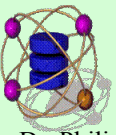
Armstrong's Axioms:

- if $\beta \subseteq \alpha$, then $\alpha \rightarrow \beta$ **(reflexivity)** (e.g., if α = AB then AB $\rightarrow$ A and AB $\rightarrow$ B)
  (these are the *trivial* dependencies)
- if $\alpha \rightarrow \beta$, then $\gamma\ \alpha \rightarrow \beta$ **(augmentation)**
  and $\gamma\ \alpha \rightarrow \gamma\ \beta$
- if $\alpha \rightarrow \beta$, and $\beta \rightarrow \gamma$, then $\alpha \rightarrow \gamma$ **(transitivity)**

Additional Rules

Additional rules implied by Armstrong's Axioms:

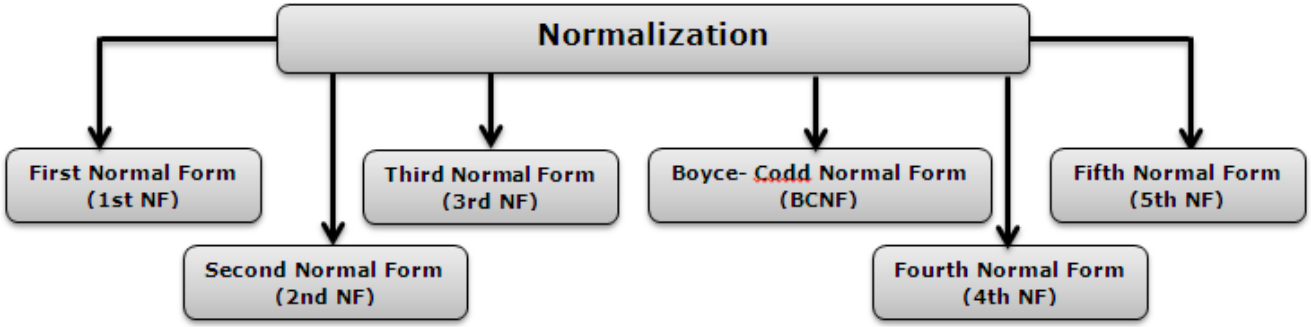
- if $\alpha \rightarrow \beta$ holds *a*nd $\alpha \rightarrow \gamma$ holds, then $\alpha \rightarrow \beta\,\gamma$ holds **(union)**
- if $\alpha \rightarrow \beta\,\gamma$ holds, then $\alpha \rightarrow \beta$ holds and $\alpha \rightarrow \gamma$ holds **(decomposition)**
- if $\alpha \rightarrow \beta$ holds *a*nd $\gamma\,\beta \rightarrow \delta$ holds, then $\alpha\,\gamma \rightarrow \delta$ holds **(pseudotransitivity)**

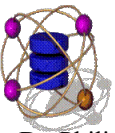


1NF

1NF - Relation R is in 1NF if:

- All data stored at the intersection of a row (tuple) and column (attribute) must be atomic (i.e., it can't be decomposed into multiple values). See next pages for examples.
- A table (Relation) must not contain any repeating columns (attributes). See next pages for examples.



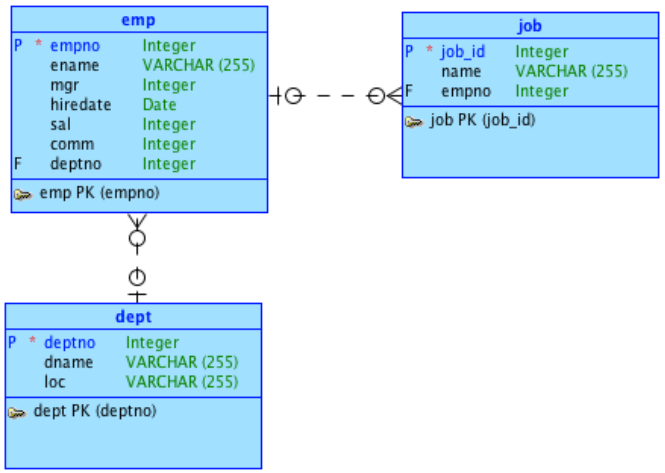


1NF – All values must be Atomic

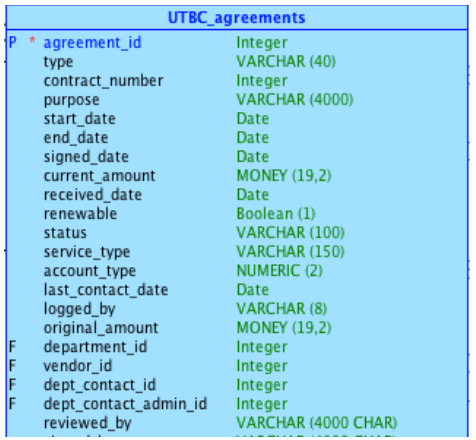
Badly Designed Table

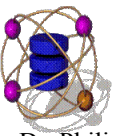
EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
7369	SMITH	CLERK, Another Job	7902	17-DEC-80	824		20
7499	ALLEN	SALESMAN, Another Job	7698	20-FEB-81	1600	300	30
7521	WARD	SALESMAN, Another Job	7698	22-FEB-81	1250	500	30
7566	JONES	MANAGER, Another Job	7839	02-APR-81	2975		20
7654	MARTIN	SALESMAN, Another Job	7698	28-SEP-81	1250	1400	30
7698	BLAKE	MANAGER, Another Job	7839	01-MAY-81	2850		30
7782	CLARK	MANAGER, Another Job	7839	09-JUN-81	2450		10
7788	SCOTT	ANALYST, Another Job	7566	09-DEC-82	3000		20
7839	KING	none, Another Job		17-NOV-81	5000		10
7844	TURNER	SALESMAN, Another Job	7698	08-SEP-81	1500		30
7876	ADAMS	CLERK, Another Job	7788	12-JAN-83	1100		20
7900	JAMES	CLERK, Another Job	7698	03-DEC-81	950		30
7902	FORD	ANALYST, Another Job	7566	03-DEC-81	3000		20
7934	MILLER	CLERK, Another Job	7782	23-JAN-82	1300		10

The Fix



Or, in some cases, create a table with a "type" column that represents a hierarchy as in the Business Contracts "Contact" table → .



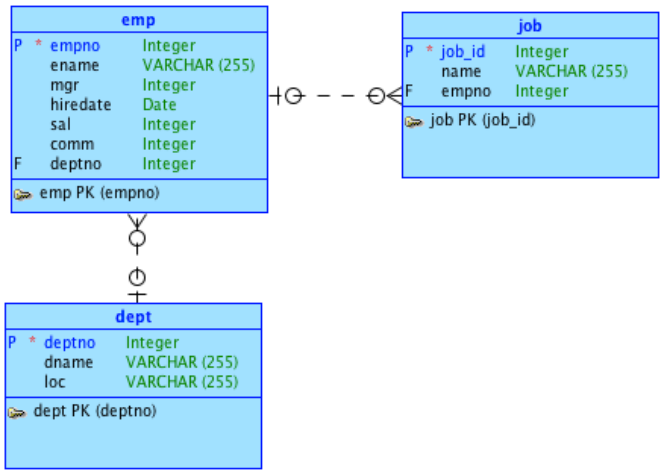


1NF – No Repeating Columns (Attributes)

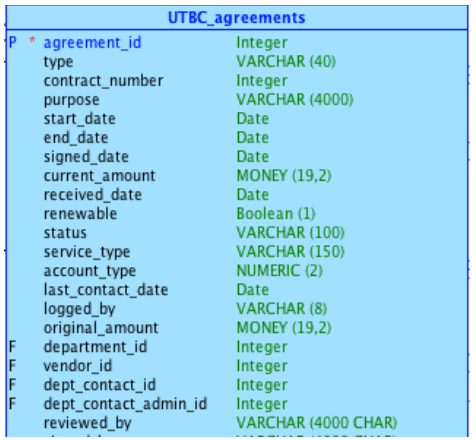
Badly Designed Table

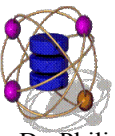
EMPNO	ENAME	JOB	JOB2	MGR	HIREDATE	SAL	COMM	DEPTNO
7369	SMITH	CLERK	Another Job	7902	17-DEC-80	824		20
7499	ALLEN	SALESMAN	Another Job	7698	20-FEB-81	1600	300	30
7521	WARD	SALESMAN	Another Job	7698	22-FEB-81	1250	500	30
7566	JONES	MANAGER	Another Job	7839	02-APR-81	2975		20
7654	MARTIN	SALESMAN	Another Job	7698	28-SEP-81	1250	1400	30
7698	BLAKE	MANAGER	Another Job	7839	01-MAY-81	2850		30
7782	CLARK	MANAGER	Another Job	7839	09-JUN-81	2450		10
7788	SCOTT	ANALYST	Another Job	7566	09-DEC-82	3000		20
7839	KING	none	Another Job		17-NOV-81	5000		10
7844	TURNER	SALESMAN	Another Job	7698	08-SEP-81	1500		30
7876	ADAMS	CLERK	Another Job	7788	12-JAN-83	1100		20
7900	JAMES	CLERK	Another Job	7698	03-DEC-81	950		30
7902	FORD	ANALYST	Another Job	7566	03-DEC-81	3000		20
7934	MILLER	CLERK	Another Job	7782	23-JAN-82	1300		

The Fix



Or, in some cases, create a table with a "type" column that represents a hierarchy as in the Business Contracts "Contact" table → .





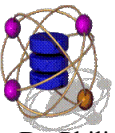
2NF

2NF - Relation R is in 2NF if:

R is in 1NF and

- there is no nonprime attribute (i.e., an attribute that's not part of *any* candidate key) that is dependent on any ***part*** of the candidate key
- *(no nonprime attribute is functionally determined by a prime attribute)*
- in other words, each nonprime attribute in relation R is *fully* dependent upon every candidate key

Note: If relation R has no compound keys, R is automatically in 2NF

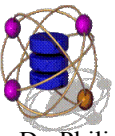


A Badly Designed (Non-2NF) Table*

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	DNAME	LOC
7369	SMITH	CLERK	7902	17-DEC-80	824		20	RESEARCH	DALLAS
7369	SMITH	CLERK	7902	17-DEC-80	824		30	SALES	CHICAGO
7499	ALLEN	SALESMAN	7698	20-FEB-81	1600	300	10	ACCOUNTING	NEW YORK
7499	ALLEN	SALESMAN	7698	20-FEB-81	1600	300	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	22-FEB-81	1250	500	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	22-FEB-81	1250	500	10	ACCOUNTING	NEW YORK
7566	JONES	MANAGER	7839	02-APR-81	2975		30	SALES	CHICAGO
7566	JONES	MANAGER	7839	02-APR-81	2975		20	RESEARCH	DALLAS
7654	MARTIN	SALESMAN	7698	28-SEP-81	1250	1400	30	SALES	CHICAGO
7654	MARTIN	SALESMAN	7698	28-SEP-81	1250	1400	10	ACCOUNTING	NEW YORK
7698	BLAKE	MANAGER	7839	01-MAY-81	2850		30	SALES	CHICAGO
7698	BLAKE	MANAGER	7839	01-MAY-81	2850		10	ACCOUNTING	NEW YORK
7782	CLARK	MANAGER	7839	09-JUN-81	2450		10	ACCOUNTING	NEW YORK
7782	CLARK	MANAGER	7839	09-JUN-81	2450		20	RESEARCH	DALLAS
7788	SCOTT	ANALYST	7566	09-DEC-82	3000		20	RESEARCH	DALLAS
7788	SCOTT	ANALYST	7566	09-DEC-82	3000		30	SALES	CHICAGO
7839	KING	none		17-NOV-81	5000		10	ACCOUNTING	NEW YORK
7839	KING	none		17-NOV-81	5000		20	RESEARCH	DALLAS
7844	TURNER	SALESMAN	7698	08-SEP-81	1500		30	SALES	CHICAGO
7844	TURNER	SALESMAN	7698	08-SEP-81	1500		10	ACCOUNTING	NEW YORK
7876	ADAMS	CLERK	7788	12-JAN-83	1100		20	RESEARCH	DALLAS
7876	ADAMS	CLERK	7788	12-JAN-83	1100		30	SALES	CHICAGO
7900	JAMES	CLERK	7698	03-DEC-81	950		30	SALES	CHICAGO
7900	JAMES	CLERK	7698	03-DEC-81	950		10	ACCOUNTING	NEW YORK
7902	FORD	ANALYST	7566	03-DEC-81	3000		20	RESEARCH	DALLAS
7902	FORD	ANALYST	7566	03-DEC-81	3000		30	SALES	CHICAGO
7934	MILLER	CLERK	7782	23-JAN-82	1300		20	RESEARCH	DALLAS
7934	MILLER	CLERK	7782	23-JAN-82	1300		10	ACCOUNTING	NEW YORK

*This table was create with the following SQL:

```
insert into emp_dept (select t.*, dname, loc from (select EMPNO, ENAME, JOB, MGR,HIREDATE, SAL, COMM,
case
when deptno = 10 then 20
when deptno = 20 then 30
when deptno = 30 then 10
end deptno from emp) t, dept d where t.deptno = d.deptno)
```

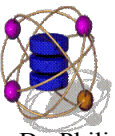


Designing Databases Using Normalization

2NF – <u>A good way to remember this</u>.

If you have a SINGLE table that is a mapping from the following object model (instead of the normal three tables),

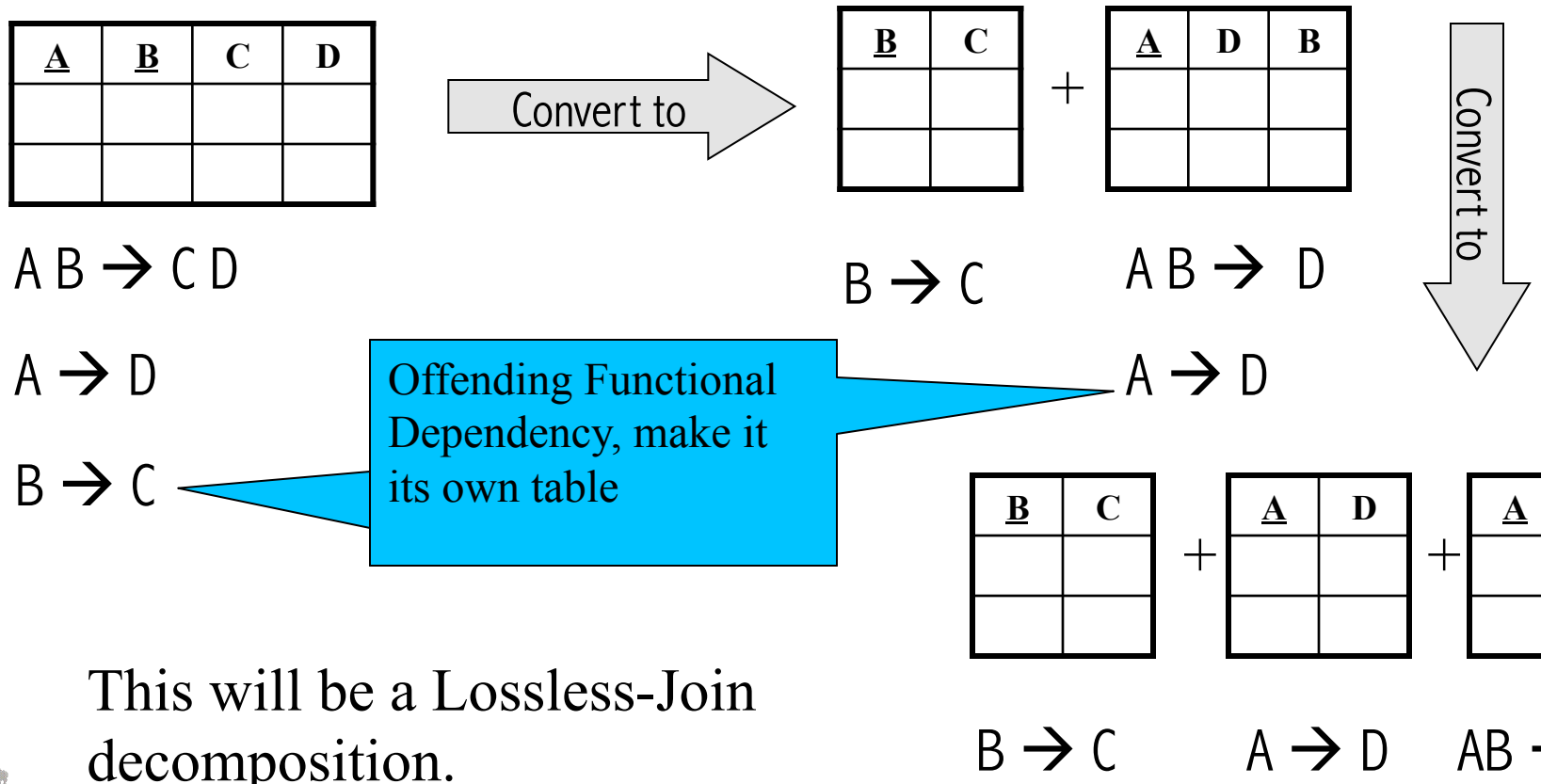
the single table will not be in 2NF.



Designing Databases Using Normalization

The problem is that some non-key attributes are functionally dependent on PART of the key (this is not allowed for 2NF).

Decomposition into 2NF

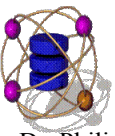


This will be a Lossless-Join decomposition.

3NF

3NF - Relation R is in 3NF if:

- R is in 2NF and
- No nonprime attribute functionally determines any other nonprime attribute



A Badly Designed Table*

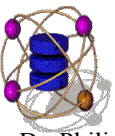
EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO	DNAME	LOC
7369	SMITH	CLERK	7902	17-DEC-80	824		20	RESEARCH	DALLAS
7499	ALLEN	SALESMAN	7698	20-FEB-81	1600	300	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	22-FEB-81	1250	500	30	SALES	CHICAGO
7566	JONES	MANAGER	7839	02-APR-81	2975		20	RESEARCH	DALLAS
7654	MARTIN	SALESMAN	7698	28-SEP-81	1250	1400	30	SALES	CHICAGO
7698	BLAKE	MANAGER	7839	01-MAY-81	2850		30	SALES	CHICAGO
7782	CLARK	MANAGER	7839	09-JUN-81	2450		10	ACCOUNTING	NEW YORK
7788	SCOTT	ANALYST	7566	09-DEC-82	3000		20	RESEARCH	DALLAS
7839	KING	none		17-NOV-81	5000		10	ACCOUNTING	NEW YORK
7844	TURNER	SALESMAN	7698	08-SEP-81	1500		30	SALES	CHICAGO
7876	ADAMS	CLERK	7788	12-JAN-83	1100		20	RESEARCH	DALLAS
7900	JAMES	CLERK	7698	03-DEC-81	950		30	SALES	CHICAGO
7902	FORD	ANALYST	7566	03-DEC-81	3000		20	RESEARCH	DALLAS
7934	MILLER	CLERK	7782	23-JAN-82	1300		10	ACCOUNTING	NEW YORK

This table was create with the following SQL: create table emp_dept as (select e., dname, loc from emp e, dept d where e.deptno = d.deptno)

In this table EMPNO -> ENAME JOB MGR HIREDATE SAL COMM DEPTNO DNAME LOC

and DEPTNO -> DNAME LOC

-> stands for "Funcionally Determines" which we will discuss later.

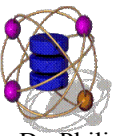


Designing Databases Using Normalization

3NF – <u>A good way to remember this</u>.

If you have a SINGLE table that is a mapping from the following object model (instead of the normal two tables),

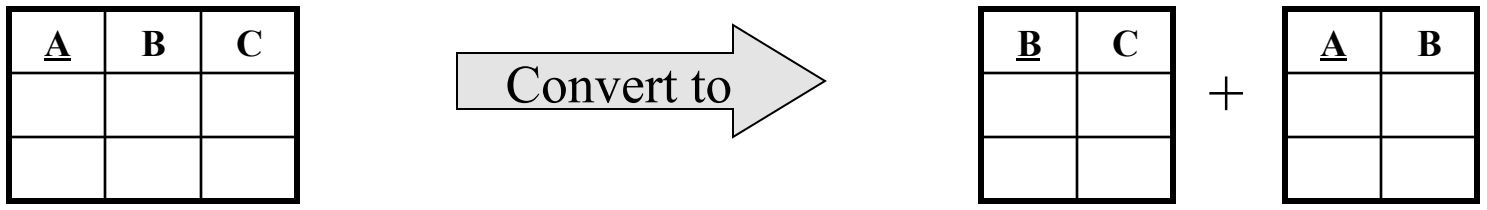
the single table will not be in 3NF.



Designing Databases Using Normalization

The problem is that some non-key attribute(s) are functionally dependent on a non-key attribute (this is not allowed for 3NF).

Decomposition into 3NF



A → B C

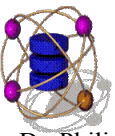
B → C

Offending Functional Dependency, make it its own table

B → C

A → B

This will be a Lossless-Join decomposition, see the next 2 pages for a description of a Lossless-Join decomposition.



Designing Databases Using Normalization

Lossless-Join Decomposition

Example of Non Lossless-Join Decomposition

Decompose empdept (create table empdept as select * from emp natural join dept) into 2 tables:

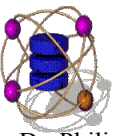
create table emp_dept1 as

select empno, ename, job, mgr, hiredate, sal, comm from emp_dept

create table emp_dept2 as select job, deptno, dname, loc from emp_dept

Then try to recreate empdept by issuing the following query

select * from emp_dept1 natural join emp_dept2



Designing Databases Using Normalization

Lossless-Join Decomposition

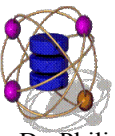
Test for Lossless-Join Decomposition

Let R be a relation that you want to decompose into R1 and R2.

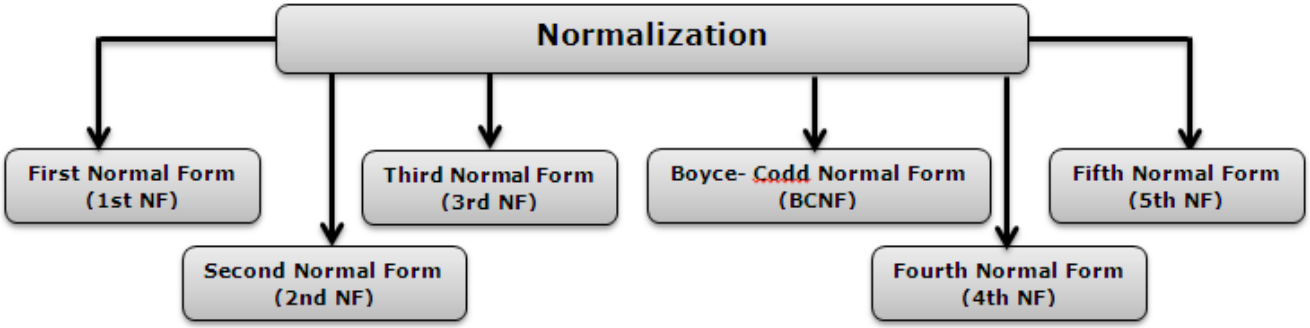
The decomposition is Lossless if

$R1 \cap R2 \rightarrow R1$ or $R1 \cap R2 \rightarrow R2$

In other words, if $R1 \cap R2$ forms a superkey of either R1 or R2



Try This



Work from left to right →

Given the Relation R=ABCDEF that's in 1NF and the following set F of Functional Dependencies,

A→BC
D→EF
E→F

find a set of Relations that are in 3NF.

You might need the following to find the Primary Key of R.

Armstrong's Axioms:

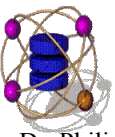
- if $\beta \subseteq \alpha$, then $\alpha \rightarrow \beta$ **(reflexivity)** (e.g., if α = AB then AB → A and AB → B)

  (these are the *trivial* dependencies)
- if $\alpha \rightarrow \beta$, then $\gamma\ \alpha \rightarrow \beta$ **(augmentation)**

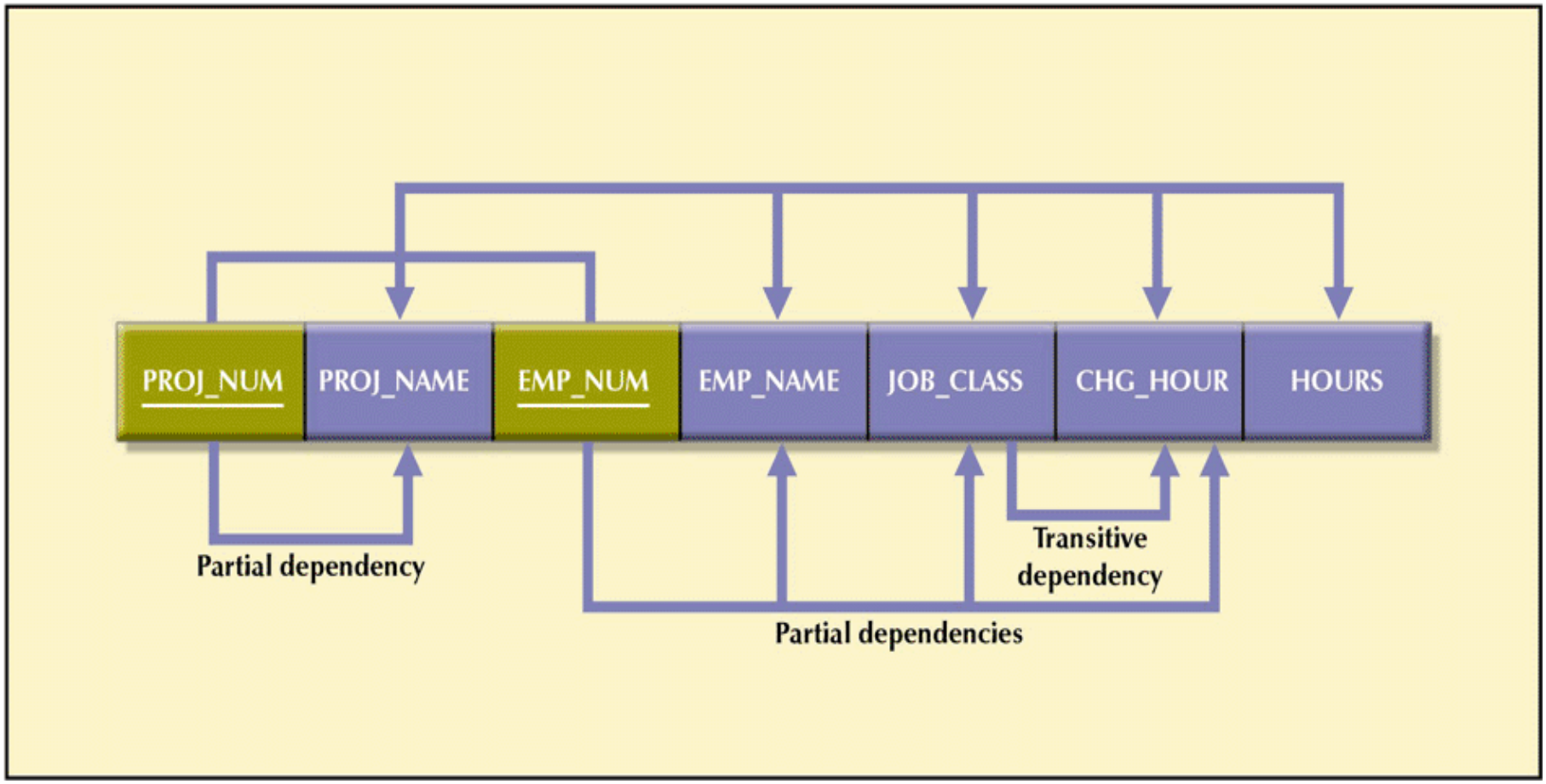
  and $\gamma\ \alpha \rightarrow \gamma\ \beta$
- if $\alpha \rightarrow \beta$, and $\beta \rightarrow \gamma$, then $\alpha \rightarrow \gamma$ **(transitivity)**

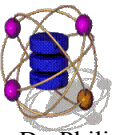
Additional rules implied by Armstrong's Axioms:

- if $\alpha \rightarrow \beta$ holds *and* $\alpha \rightarrow \gamma$ holds, then $\alpha \rightarrow \beta\ \gamma$ holds **(union)**
- if $\alpha \rightarrow \beta\ \gamma$ holds, then $\alpha \rightarrow \beta$ holds and $\alpha \rightarrow \gamma$ holds **(decomposition)**
- if $\alpha \rightarrow \beta$ holds *and* $\gamma\ \beta \rightarrow \delta$ holds, then $\alpha\ \gamma \rightarrow \delta$ holds **(pseudotransitivity)**

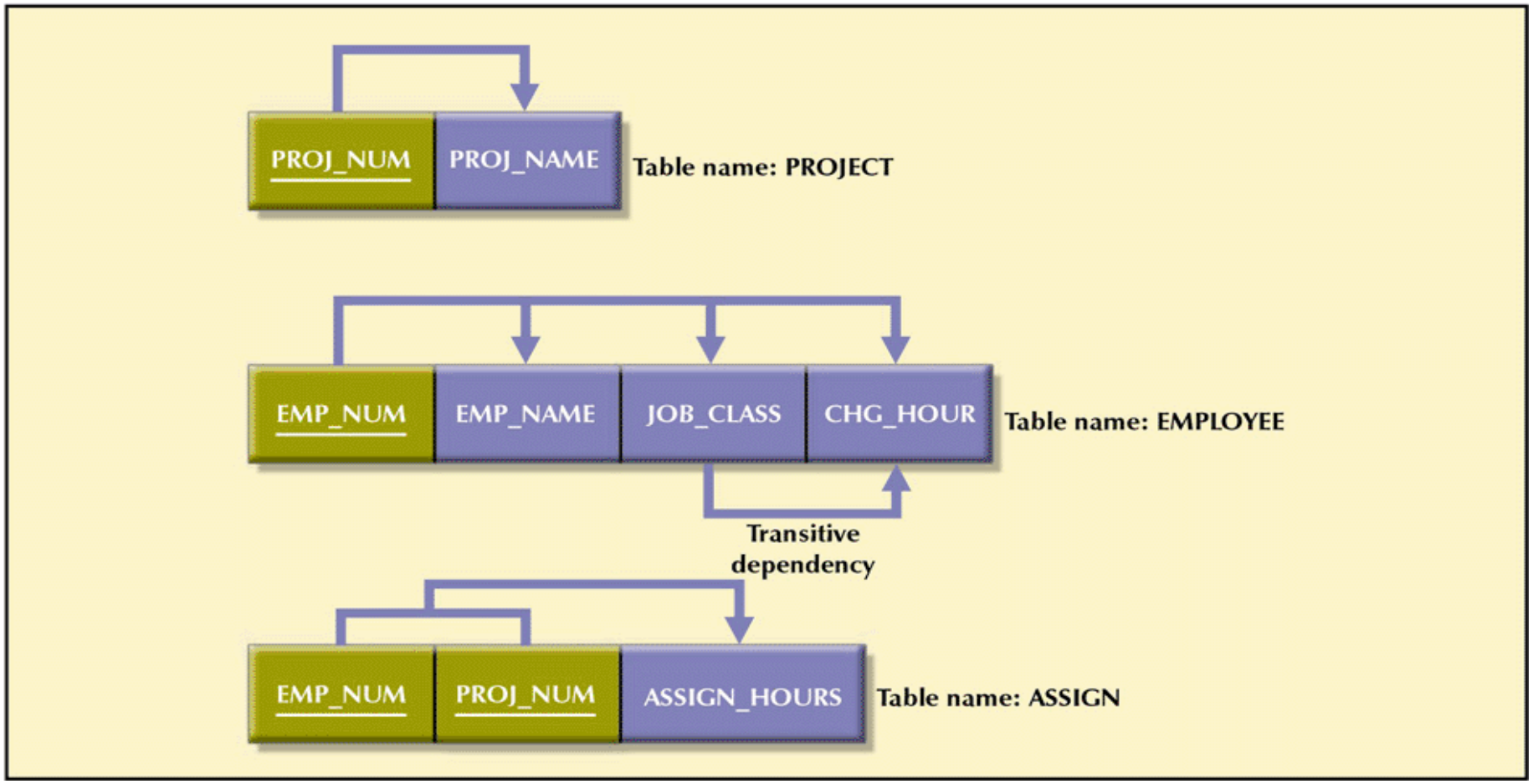


Another Example





Second Normal Form (2NF) Conversion Results



Third Normal Form (3NF) Conversion Results