Unit 7
Logical Database Design
With Normalization

Logical Database Design

- We are given a set of tables specifying the database
  - The base tables, which probably are the community (conceptual) level
- They may have come from some ER diagram or from somewhere else
- We will need to examine whether the specific choice of tables is good for
  - Storing the information needed
  - Enforcing constraints
  - Avoiding anomalies, such as redundancies
- If there are problems to address, we may want to restructure the database, of course not losing any information
- Let us quickly review an example from "long time ago"

`

R	Name	SSN	DOB	Grade	Salary
	A	121	2367	2	80
	A	132	3678	3	70
	B	101	3498	4	70
	C	106	2987	2	80

- Business rule (one among several):
  - The value of Salary is determined only by the value of Grade
- Comment:
  - We keep track of the various Grades for more than just computing salaries, though we do not show it
  - For instance, DOB and Grade together determine the number of vacation days, which may therefore be different for SSN 121 and 106

Anomalies

Name	<u>SSN</u>	DOB	Grade	Salary
A	121	2367	2	80
A	132	3678	3	70
B	101	3498	4	70
C	106	2987	2	80

- “Grade = 2 implies Salary = 80” is written twice
- There are additional problems with this design.
  - We are unable to store the salary structure for a Grade that does not currently exist for any employee.
  - For example, we cannot store that Grade = 1 implies Salary = 90
  - For example, if employee with SSN = 132 leaves, we forget which Salary should be paid to employee with Grade = 3
  - We could perhaps invent a fake employee with such a Grade and such a Salary, but this brings up additional problems, e.g.,

    What is the SSN of such a fake employee? It cannot be NULL as SSN is the primary key

Better Representation Of Information

- The problem can be solved by replacing

R	Name	SSN	DOB	Grade	Salary
	A	121	2367	2	80
	A	132	3678	3	70
	B	101	3498	4	70
	C	106	2987	2	80

by two tables

S	Name	SSN	DOB	Grade
	A	121	2367	2
	A	132	3678	3
	B	101	3498	4
	C	106	2987	2

T	Grade	Salary
	2	80
	3	70
	4	70

Decomposition

- SELECT INTO S
  Name, SSN, DOB, Grade
  FROM R;

- SELECT INTO T
  Grade, Salary
  FROM R;

Better Representation Of Information

- And now we can
  - Store "Grade = 3 implies Salary = 70", even after the last employee with this Grade leaves
  - Store "Grade = 1 implies Salary = 90", planning for hiring employees with Grade = 1, while we do not yet have any employees with this Grade

S	Name	SSN	DOB	Grade
	A	121	2367	2
	B	101	3498	4
	C	106	2987	2

T	Grade	Salary
	1	90
	2	80
	3	70
	4	70

No Information Was Lost

- Given S and T, we can reconstruct R using ***natural join***

S	Name	<u>SSN</u>	DOB	Grade
	A	121	2367	2
	A	132	3678	3
	B	101	3498	4
	C	106	2987	2

T	<u>Grade</u>	Salary
	2	80
	3	70
	4	70

```
SELECT INTO R
Name, SSN, DOB, S.Grade AS Grade, Salary
FROM T, S
WHERE T.Grade = S.Grade;
```

R	Name	<u>SSN</u>	DOB	Grade	Salary
	A	121	2367	2	80
	A	132	3678	3	70
	B	101	3498	4	70
	C	106	2987	2	80

Natural Join

- Given several tables, say R_1, R_2, …, R_n, their ***natural join*** is computed using the following "template":

  SELECT INTO R
  one copy of each column name
  FROM R_1, R_2, …, R_n
  WHERE equal-named columns have to be equal

- The intuition is that R was "decomposed" into R_1, R_2, …,R_n by appropriate SELECT statements, and now we are putting it back together

Natural Join And Lossless Join Decomposition

- ***Natural Join*** is:
  - Cartesian join with condition of equality on corresponding columns
  - Only one copy of each column is kept
- ***"Lossless join decomposition"*** is another term for information not being lost, that is we can reconstruct the original table by "combining" information from the two new tables by means of natural join
- This does not necessarily always hold
- We will have more material about this later
- Here we just observe that our decomposition satisfied this condition at least in our example

Mathematical Notation For Natural Join (We Will Use Sparingly)

- There is a special mathematical symbol for natural join
- It is not part of SQL, of course, which only allows standard ANSI font

- In mathematical, relational algebra notation, natural join of two tables is denoted by $\bowtie$ (this symbol appears only in special mathematical fonts, so sometimes ∞ may be used instead)

- So we have: $R = S \bowtie T$

- It is used when "corresponding columns" means "equal columns"

Revisiting The Problem

- Let us look at

R	Name	<u>SSN</u>	DOB	Grade	Salary
	A	121	2367	2	80
	A	132	3678	3	70
	B	101	3498	4	70
	C	106	2987	2	80
	A	132	3678	3	70
	B	101	3498	4	70

- The problem is ***not*** that there are duplicate rows
- The problem is the same as before, business rule assigning Salary to Grade is written a number of times

- So how can we “generalize” the problem?

Stating The Problem In General

R	Name	SSN	DOB	Grade	Salary
	A	121	2367	2	80
	A	132	3678	3	70
	B	101	3498	4	70
	C	106	2987	2	80
	A	132	3678	3	70
	B	101	3498	4	70

- We have a problem whenever we have two sets of columns X and Y (here X is just Grade and Y is just Salary), such that
  1. **X does not contain a key either primary or unique** (thus there could be several/many **non-identical** rows with the same value of X)
  2. **Whenever two rows are equal on X, they must be equal on Y**
- Why a problem: the business rule specifying how X “forces” Y is “embedded” in different rows and therefore
  - Inherently written redundantly
  - Cannot be stored by itself

What Did We Do?
Think X = Grade And Y = Salary

- We had a table

U	X	V	Y	W

- We replaced this one table by two tables

U	X	V	W

X	Y

Logical Database Design

- We will discuss techniques for dealing with the above issues
- Formally, we will study ***normalization*** (decompositions as in the above example) and ***normal forms*** (forms for relation specifying some “niceness” conditions)
- There will be three very important issues of interest:
  - Removal of redundancies
  - Lossless-join decompositions
  - Preservation of dependencies
- We will learn the material mostly through comprehensive examples
- But everything will be precisely defined
- Algorithms will be fully and precisely given in the material
- Some of this will be part of the Advanced part of this Unit

Several Passes On The Material

- Practitioners do it (mostly) differently than the way researchers/academics like to do

- **Pass 1**: We focus on how IT practitioners do it or at least like to talk about it

  Ad-hoc treatment, but good for building intuition and having common language and concepts with IT people

- **Pass 2**: We focus on how computer scientists like to do or at least can do it this way if they want to

  Good for actually using algorithms that guarantee correct results

The Topic Is Normalization And Normal Forms

- Normalization deals with “reorganizing” a relational database by, generally, breaking up tables (relations) to remove various anomalies
- We start with the way practitioners think about it (as we have just said)
- We will proceed by means of a simple example, which is rich enough to understand what the problems are and how to think about fixing them
- It is important (in this context) to understand what the various normal forms are (they may ask you this during a job interview!)

Normal Forms

- A normal form applies to a table/relation schema, not to the whole database schema
- So the question is individually asked about a table: is it of some specific desirable normal form?
- The ones you need to know about in increasing order of "quality" and complexity:
  - First Normal Form (**1NF**); it essentially states that we have a table/ relation
  - Second Normal Form (**2NF**); intermediate form in some algorithms
  - Third Normal Form (**3NF**); very important; a final form
  - Boyce-Codd Normal Form (**BCNF**); very important in theory (but less used in practice and we will understand why); a final form
  - Fourth Normal Form (**4NF**); a final form but generally what is good about it beyond previous normal forms is easily obtained without formal treatment
- There are additional ones, which are more esoteric, and which we will not cover

Our Example

- We will deal with a very small fragment of a database dealing with a university
- We will make some assumptions in order to focus on the points that we need to learn

- We will identify people completely by their first names, which will be like Social Security Numbers
  - That is, whenever we see a particular first name more than once, such as Fang or Allan, this will always refer to the same person: there is only one Fang in the university, etc.

Our New Example

- We are looking at a single table in our database
- It has the following columns
  - S, which is a Student
  - B, which is the Birth Year of the Student
  - C, which is a Course that the student took
  - T, which is the Teacher who taught the Course the Student took
  - F, which is the credits that the Teacher is given for having the Student take the course
- We will start with something that is not even a relation (Note this is similar to Employees having Children in Unit 2; a Student may have any number of (Course,Teacher,Fee) values

	S	B	C	T	F	C	T	F
	Fang	1990	DB	Zvi	1	OS	Allan	2
	John	1980	OS	Allan	2	PL	Marsha	4
	Mary	1990	PL	Vijay	1			

Alternative Depiction

- Instead of

	S	B	C	T	F	C	T	F
	Fang	1990	DB	Zvi	1	OS	Allan	2
	John	1980	OS	Allan	2	PL	Marsha	4
	Mary	1990	PL	Vijay	1			

you may see the above written as

	S	B	C	T	F
	Fang	1990	DB	Zvi	1
			OS	Allan	2
	John	1980	OS	Allan	2
			PL	Marsha	4
	Mary	1990	PL	Vijay	1

First Normal Form: ***A Table With Fixed Number Of Column***

- This ***was not*** a relation, because we are told that each Student may have taken any number of Courses
- Therefore, the number of columns is not fixed/bounded
- It is easy to make this a relation, getting

R	S	B	C	T	F
	Fang	1990	DB	Zvi	1
	John	1980	OS	Allan	2
	Mary	1990	PL	Vijay	1
	Fang	1990	OS	Allan	2
	John	1980	PL	Marsha	4

- Formally, we have a relation in ***First Normal Form (1NF)***, this means that there are no repeating groups and the number of columns is fixed: in other words this is a relation, nothing new, defined for historical reasons
  - There are some variations to this definition, but we use this one

Our Business Rules (Constraints)

- Our enterprise has certain ***business rules***
- We are told the following business rules
  1. A student can have only one birth year
  2. A teacher has to be given the same number of credits for every student he/she teaches.
  3. A teacher can teach only one course (perhaps at different times, different offerings, etc., but never another course)
  4. A student can take any specific course from one teacher only (or not at all)
- This means, that we are ***guaranteed*** that the information will always obey these business rules, as in the example

R	S	B	C	T	F
	Fang	1990	DB	Zvi	1
	John	1980	OS	Allan	2
	Mary	1990	PL	Vijay	1
	Fang	1990	OS	Allan	2
	John	1980	PL	Marsha	4

Functional Dependencies (Abbreviation: FDs)

- These rules can be formally described using ***functional dependencies***
- We will ignore NULLS
- Let P and Q be sets of columns, then:

  P ***functionally determines*** Q, written P → Q

  if and only if

  any two rows that are equal on (all the attributes in) P must be equal on (all the attributes in) Q
- In simpler terms, less formally, but really the same, it means that:

  If a value of P is specified, it "forces" some (specific) value of Q; in other words: Q is a function of P

  Note: this does not mean Q can be computed from P (only)
- In our old example we looked at Grade → Salary

Our Given Functional Dependencies

R	S	B	C	T	F
	Fang	1990	DB	Zvi	1
	John	1980	OS	Allan	2
	Mary	1990	PL	Vijay	1
	Fang	1990	OS	Allan	2
	John	1980	PL	Marsha	4

- Our rules
  1. A student can have only one birth year: $S \rightarrow B$
  2. A teacher has to be given same credits for every student he/she teaches : $T \rightarrow F$
  3. A teacher can teach only one course (perhaps at different times, different offerings, etc, but never another course) : $T \rightarrow C$
  4. A student can take a course from one teacher only: $SC \rightarrow T$

Possible Primary Key

- Our rules: S → B, T → F, T → C, SC → T
- ST possible primary key, because given ST
  1. S determines B
  2. T determines F
  3. T determines C
- A part of ST is not sufficient
  1. From S, we cannot get T, C, or F
  2. From T, we cannot get S or B

R	<u>S</u>	B	C	<u>T</u>	F
	Fang	1990	DB	Zvi	1
	John	1980	OS	Allan	2
	Mary	1990	PL	Vijay	1
	Fang	1990	OS	Allan	2
	John	1980	PL	Marsha	4

Possible Primary Key

- Our rules: $S \rightarrow B$, $T \rightarrow F$, $T \rightarrow C$, $SC \rightarrow T$
- SC possible primary key, because given SC
  1. S determines B
  2. SC determines T
  3. T determines F (we can now use T to determine F because of rule 2)
- A part of SC is not sufficient
  1. From S, we cannot get T, C, or F
  2. From C, we cannot get B, S, T, or F

R	<u>S</u>	B	<u>C</u>	T	F
	Fang	1990	DB	Zvi	1
	John	1980	OS	Allan	2
	Mary	1990	PL	Vijay	1
	Fang	1990	OS	Allan	2
	John	1980	PL	Marsha	4

Possible Primary Keys

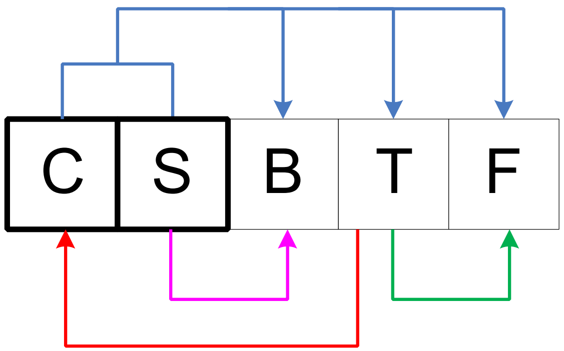
- Our rules: S → B, T → F, T → C, SC → T
- Because ST can serve as primary key, in effect:
  - ST → SBCTF
  - This is sometimes just written as ST → BCF, since always ST → ST (columns determine themselves)
- Because SC can serve as primary key, in effect:
  - SC → SBCTF
  - This is sometimes just written as SC → BTF, since always SC → SC (columns determine themselves)

We Choose The Primary Key

- We choose SC as ***the primary key***
- This choice is arbitrary, but perhaps it is more intuitively justifiable than ST
- For the time being, we ignore the other possible primary key (ST)

R	S	B	C	T	F
	Fang	1990	DB	Zvi	1
	John	1980	OS	Allan	2
	Mary	1990	PL	Vijay	1
	Fang	1990	OS	Allan	2
	John	1980	PL	Marsha	4

Drawing Functional Dependencies



- Each column in a box
- Our key (there could be more than one) is chosen to be the primary key and its boxes have thick borders and it is stored in the left part of the rectangle
- Above the boxes, we have functional dependencies ***"from the full key"*** (this is actually not necessary to draw)
- Below the boxes, we have functional dependencies ***"not from the full key"***
- Colors of lines are not important, but good for explaining

Anomalies

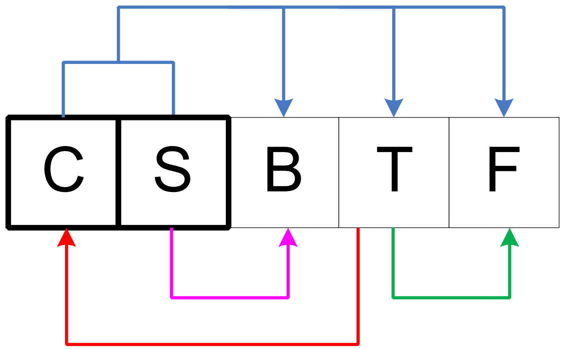
- These “not from the full key” dependencies cause the design to be bad
  - Inability to store important information
  - Redundancies
- Imagine a new Student appears who has not yet registered for a course
  - This S has a specific B, but this cannot be stored in the table as we do not have a value of C yet, and the attributes of the primary key cannot be NULL
- Imagine that Mary withdrew from the only Course she has
  - We have no way of storing her B
- Imagine that we “erase” the value of C in the row stating that Fang was taught by Allan
  - We will know that this was OS, as John was taught OS by Allan, and every teacher teaches only one subject, so we had a redundancy; and whenever there is a redundancy, there is potential for inconsistency

Anomalies

- The way to handle the problems is to replace a table with other equivalent tables that do not have these problems

- Implicitly we think as if the table had only one key (we are not paying attention to keys that are not primary)
- In fact, as we have seen, there is one more key, we just do not think about it (at least for now)

Review Of Our "Not From The Full Key" Functional Dependencies



- S → B: **partial**; called partial because the left hand side is only a proper part/subset of the key
- T → F: **transitive**; called transitive because as T is outside the key, it of course depends on the key, so we have CS → T and T → F; and therefore CS → F

  Actually, it is more correct (and sometimes done) to say that CS → F is a transitive dependency because it can be decomposed into SC → T and T → F, and then derived by transitivity
- T → C: **into the key** (from outside the key)

Classification Of The Dependencies: Warning

- Practitioners ***do not*** use consistent definitions for these
- I picked one set of definitions to use here

- We will later have formal machinery to discuss this

- Wikipedia seems to be OK, but other sources of material on the web are frequently wrong (including very respectable ones!)

Redundancies In Our Example

	S	B	C	T	F
	Fang	1990	DB	Zvi	1
	John	1980	OS	Allan	2
	Mary	1990	PL	Vijay	1
	Fang	?	?	Allan	?
	John	?	PL	Marsha	4

- What could be "recovered" if somebody covered up values (the values are not NULL)?
- All of the empty slots, marked here with "?"

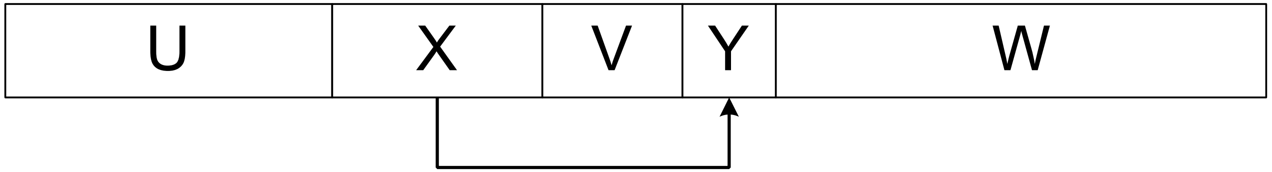
Our Business Rules Have A Clean Format

- Whoever gave them to us, understood the application very well
- The procedure we describe next assumes rules in such a clean format
- Later we will learn how to “clean” business rules without understanding the application

- Computer scientists do not assume that they understand the application or that the business rules are clean, so they use algorithmic techniques to clean up business rules
- And computer scientists prefer to use algorithms and rely less on intuition

A Procedure For Removing Anomalies

- Recall what we did with the example of Grade determining Salary
- In general, we will have sets of attributes: U, X, V, Y, W
- We replaced R(Name,SSN,DOB,Grade,Salary), where Grade → Salary; in the drawing "X" stands for "Grade" and "Y" stands for "Salary"



by two tables S(Name,SSN,DOB,Grade) and T(Grade,Salary)

- We will do the same thing, dealing with one anomaly at a time

A Procedure For Removing Anomalies

- While replacing

U	X	V	Y	W

by two tables

U	X	V	W

X	Y

- We do this if Y does not overlap (or is a part of) primary key

- We do not want to “lose” the primary key of the table UXVW, and if Y is not part of primary key of UXVYW, the primary key of UXVYW is part of UXVW and therefore it is a primary key there (a small proof is omitted)

Incorrect Decomposition (Not A Lossless Join Decomposition)

- Assume we replaced

R	Name	SSN	DOB	Grade	Salary
	A	121	2367	2	80
	A	132	3678	3	70
	B	101	3498	4	70
	C	106	2987	2	80

with two tables (note "Y" in the previous slide), which is SSN was actually the key, therefore we should not do it), without indicating the key for S to simplify the example

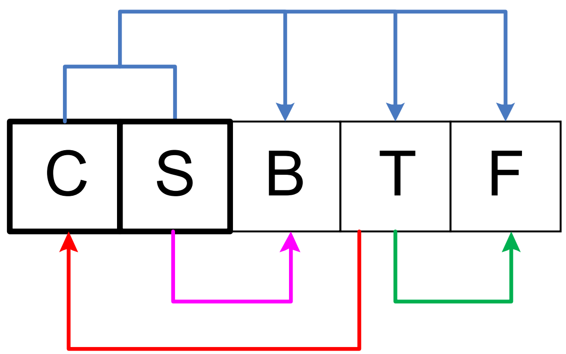
S	Name	DOB	Grade	Salary
	A	2367	2	80
	A	3678	3	70
	B	3498	4	70
	C	2987	2	80

T	SSN	Salary
	121	80
	132	70
	101	70
	106	80

- We cannot answer the question what is the Name for SSN = 121 (we lost information), so cannot decompose like this

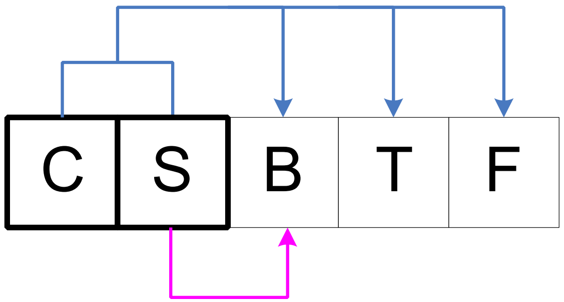
Our Example Again

	<u>S</u>	B	<u>C</u>	T	F
	Fang	1990	DB	Zvi	1
	John	1980	OS	Allan	2
	Mary	1990	PL	Vijay	1
	Fang	1990	OS	Allan	2
	John	1980	PL	Marsha	4



Partial Dependency: S → B

	<u>S</u>	B	<u>C</u>	T	F
	Fang	1990	DB	Zvi	1
	John	1980	OS	Allan	2
	Mary	1990	PL	Vijay	1
	Fang	1990	OS	Allan	2
	John	1980	PL	Marsha	4



Decomposition

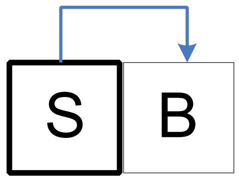
S	B	C	T	F
Fang	1990	DB	Zvi	1
John	1980	OS	Allan	2
Mary	1990	PL	Vijay	1
Fang	1990	OS	Allan	2
John	1980	PL	Marsha	4

S	B
Fang	1990
John	1980
Mary	1990
Fang	1990
John	1980

S	C	T	F
Fang	DB	Zvi	1
John	OS	Allan	2
Mary	PL	Vijay	1
Fang	OS	Allan	2
John	PL	Marsha	4

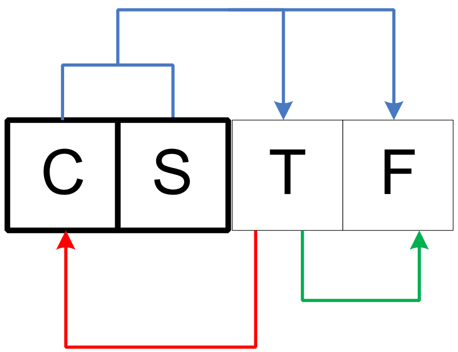
No Anomalies

	<u>**S**</u>	**B**
	Fang	1990
	John	1980
	Mary	1990
	Fang	1990
	John	1980



Some Anomalies

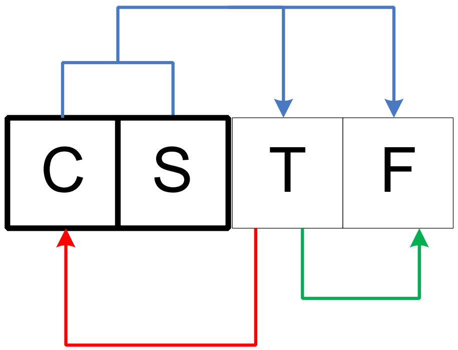
	S	C	T	F
	Fang	DB	Zvi	1
	John	OS	Allan	2
	Mary	PL	Vijay	1
	Fang	OS	Allan	2
	John	PL	Marsha	4

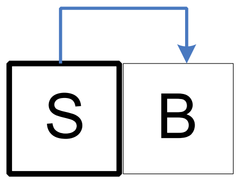


Decomposition So Far

	S	C	T	F
	Fang	DB	Zvi	1
	John	OS	Allan	2
	Mary	PL	Vijay	1
	Fang	OS	Allan	2
	John	PL	Marsha	4

	S	B
	Fang	1990
	John	1980
	Mary	1990





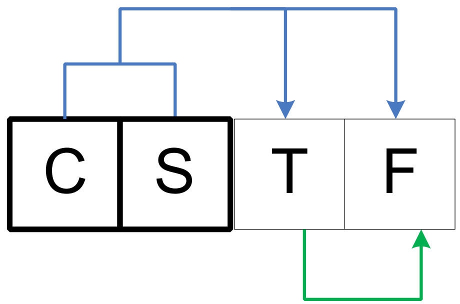
Second Normal Form: *1NF And No Partial Dependencies*

- Each of the tables in our database is in Second Normal Form
- Second Normal Form means:
  - First Normal Form
  - No partial dependencies
- The above is checked individually for each table

- Furthermore, our decomposition was a lossless join decomposition
- This means that by "combining" all the tables using the natural join, we get exactly the original table back
- This is checked "globally"; we do not discuss how this is done generally, but intuitively clearly true in our simple example

Transitive Dependency: T → F

	S	C	T	F
	Fang	DB	Zvi	1
	John	OS	Allan	2
	Mary	PL	Vijay	1
	Fang	OS	Allan	2
	John	PL	Marsha	4



Decomposition

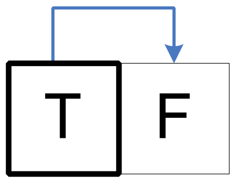
	<u>S</u>	<u>C</u>	T	F
	Fang	DB	Zvi	1
	John	OS	Allan	2
	Mary	PL	Vijay	1
	Fang	OS	Allan	2
	John	PL	Marsha	4

	<u>S</u>	<u>C</u>	T
	Fang	DB	Zvi
	John	OS	Allan
	Mary	PL	Vijay
	Fang	OS	Allan
	John	PL	Marsha

	<u>T</u>	F
	Zvi	1
	Allan	2
	Vijay	1
	Allan	2
	Marsha	4

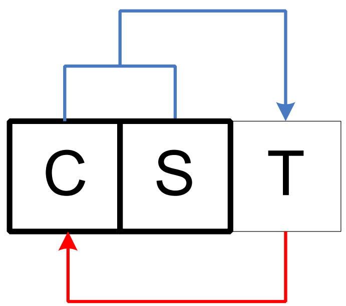
No Anomalies

	<u>I</u>	F
	Zvi	1
	Allan	2
	Vijay	1
	Allan	2
	Marsha	4



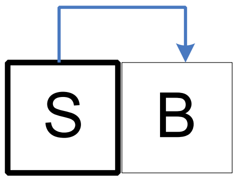
Anomalies

	S	C	T
	Fang	DB	Zvi
	John	OS	Allan
	Mary	PL	Vijay
	Fang	OS	Allan
	John	PL	Marsha

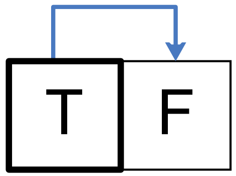


Decomposition So Far

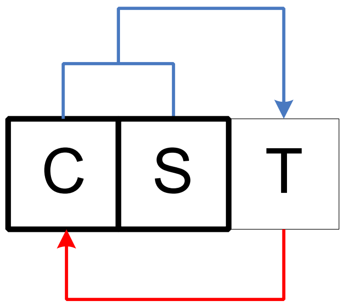
	S	B
	Fang	1990
	John	1980
	Mary	1990



	I	F
	Zvi	1
	Allan	2
	Vijay	1
	Marsha	4



	S	C	T
	Fang	DB	Zvi
	John	OS	Allan
	Mary	PL	Vijay
	Fang	OS	Allan
	John	PL	Marsha



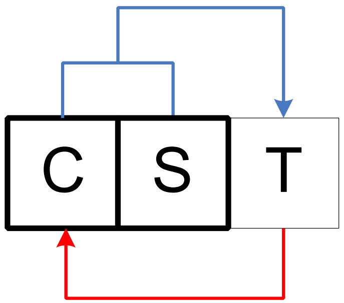
Third Normal Form:
2NF And No Transitive Dependencies

- Each of the tables in our database is in Third Normal Form
- Third Normal Form means:
  - Second Normal Form (therefore in 1NF and no partial dependencies)
  - No transitive dependencies
- The above is checked individually for each table

- Furthermore, our decomposition was a lossless join decomposition
- This means that by "combining" all the tables we get exactly the original table back
- This is checked "globally"; we do not discuss how this is done generally, but intuitively clearly true in our simple example

Anomaly

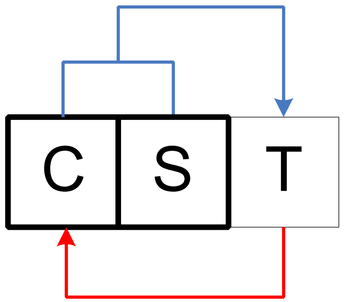
	S	C	T
	Fang	DB	Zvi
	John	OS	Allan
	Mary	PL	Vijay
	Fang	OS	Allan
	John	PL	Marsha



- We are worried about decomposing by "pulling out" C and getting CS and TC, as we are pulling out a part of the key
- But we can actually do it

An Alternative Primary Key: TS

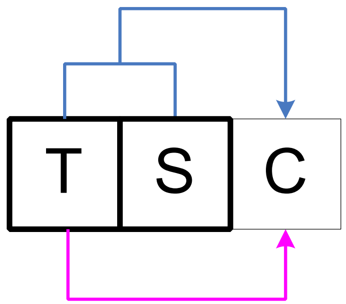
	S	C	T
	Fang	DB	Zvi
	John	OS	Allan
	Mary	PL	Vijay
	Fang	OS	Allan
	John	PL	Marsha



- Note that TS could also serve as primary key for this table SCT since by looking at the FD we have: T → C, we see that TS functionally determines everything, that is it determines all the attributes TSC
- Recall, that TS could have been chosen at the primary key of the original table

Anomaly

	S	C	I
	Fang	DB	Zvi
	John	OS	Allan
	Mary	PL	Vijay
	Fang	OS	Allan
	John	PL	Marsha



- Now our anomaly is a partial dependency, which we know how to handle

Decomposition

	S	C	I
	Fang	DB	Zvi
	John	OS	Allan
	Mary	PL	Vijay
	Fang	OS	Allan
	John	PL	Marsha

	S	I
	Fang	Zvi
	John	Allan
	Mary	Vijay
	Fang	Allan
	John	Marsha

	C	I
	DB	Zvi
	OS	Allan
	PL	Vijay
	OS	Allan
	PL	Marsha

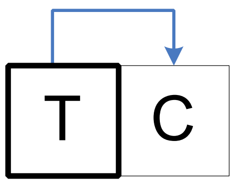
No Anomalies

	S	T
	Fang	Zvi
	John	Allan
	Mary	Vijay
	Fang	Allan
	John	Marsha

T	S

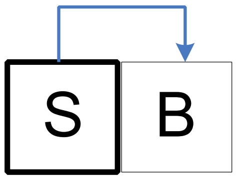
No Anomalies

	C	I
	DB	Zvi
	OS	Allan
	PL	Vijay
	OS	Allan
	PL	Marsha

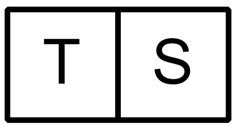


Our Decomposition

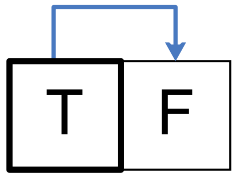
	S	B
	Fang	1990
	John	1980
	Mary	1990



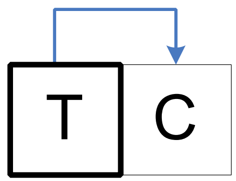
	S	I
	Fang	Zvi
	John	Allan
	Mary	Vijay
	Fang	Allan
	John	Marsha



	I	F
	Zvi	1
	Allan	2
	Vijay	1
	Marsha	4



	C	I
	DB	Zvi
	OS	Allan
	PL	Vijay
	PL	Marsha



Our Decomposition

- We can also combine tables if they have the same key and we can still maintain good properties

S	B
Fang	1990
John	1980
Mary	1990

S	B

I	F	C
Zvi	1	DB
Allan	2	OS
Vijay	1	PL
Marsha	4	PL

T	C	F

S	I
Fang	Zvi
John	Allan
Mary	Vijay
Fang	Allan
John	Marsha

T	S

Boyce-Codd Normal Form:
1NF And All Dependencies From Full Key

- Each of the tables in our database is in Boyce-Codd Normal Form
- Boyce-Codd Normal Form (BCNF) means:
  - First Normal Form
  - Every functional dependency is from a full key

    This definition is “loose.” Later, a complete, formal definition
- A table in BCNF is automatically in 3NF as no bad dependencies are possible
- The above is checked individually for each table

- Furthermore, our decomposition was a lossless join decomposition
- This means that by “combining” all the tables we get exactly the original table back
- This is checked “globally”; we do not discuss how this is done generally, but intuitively clearly true in our simple example

A New Issue: Maintaining Database Correctness And Preservation Of Dependencies

- We can understand this just by looking at the table which we decomposed last
- We will not use drawings but write the constraints that needed to be satisfied in narrative
- We will examine an update to the database and look at two scenarios
- When we have one “imperfect” 3NF table SCT
- When we have two “perfect” BCNF tables ST and CT
- We will attempt an incorrect update and see how to detect it under both scenarios

Our Tables (For The Two Cases)

- SCT satisifies: SC → T and ST →C: keys SC and ST

	<u>S</u>	<u>C</u>	T
	Fang	DB	Zvi
	John	OS	Allan
	Mary	PL	Vijay
	Fang	OS	Allan
	John	PL	Marsha

- ST does not satisfy anything: key ST
- CT satisfies T → C: key T

	<u>S</u>	<u>T</u>
	Fang	Zvi
	John	Allan
	Mary	Vijay
	Fang	Allan
	John	Marsha

	C	<u>T</u>
	DB	Zvi
	OS	Allan
	PL	Vijay
	OS	Allan
	PL	Marsha

An Insert Attempt

- A user wants to specify that now John is going to take PL from Vijay
- If we look at the database, we realize this update should not be permitted because
  - John can take PL from at most one teacher
  - John already took PL (from Marsha)
- But can the system figure this out just by checking whether FDs continue being satisified?
- Let us find out what will happen in each of the two scenarios

Scenario 1: SCT

- We maintain SCT, knowing that its keys are SC and ST

- Before the INSERT, constraints are satisfied; keys are OK

	S	C	T
	Fang	DB	Zvi
	John	OS	Allan
	Mary	PL	Vijay
	Fang	OS	Allan
	John	PL	Marsha

- After the INSERT, constraints are not satisfied; SC is no longer a key

- INSERT rejected after the constraint is checked

	S	C	T
	Fang	DB	Zvi
	John	OS	Allan
	Mary	PL	Vijay
	Fang	OS	Allan
	John	PL	Marsha
	John	PL	Vijay

Scenario 2: ST And CT

- We maintain ST, knowing that its key is ST
- We maintain CT, knowing that its key is T

- Before the INSERT, constraints are satisfied; keys are OK

	S	T		C	T
	Fang	Zvi		DB	Zvi
	John	Allan		OS	Allan
	Mary	Vijay		PL	Vijay
	Fang	Allan		OS	Allan
	John	Marsha		PL	Marsha

- After the INSERT, constraints are still satisfied; keys remain keys

	S	T		C	T
	Fang	Zvi		DB	Zvi
	John	Allan		OS	Allan
	Mary	Vijay		PL	Vijay
	Fang	Allan		OS	Allan
	John	Marsha		PL	Marsha
	John	Vijay		PL	Vijay

- But the INSERT ***must*** still be rejected

Scenario 2: What To Do?

- The INSERT must be rejected
- This bad insert cannot be discovered as bad by examining only what happens in each individual table
- The formal term for this is: ***dependencies are not preserved***

- So need to perform non-local tests to check updates for validity
- For example, take ST and CT and reconstruct SCT

A Very Important Conclusion

- Generally, normalize up to 3NF and not up to BCNF
  - So the database is not fully normalized
- Luckily, when you do this, frequently you “automatically” get BCNF
  - But not in our example, which I set up on purpose so this does not happen

Multivalued Dependencies

- To have a smaller example, we will look at this separately, not by extending our previous example
  - Otherwise, it would become too big
- In the application, we store information about Courses (C), Teachers (T), and Books (B)
- Each course has a set of books that have to be assigned during the course
- Each course has a set of teachers that are qualified to teach the course
- Each teacher, when teaching a course, has to use the set of the books that has to be assigned in the course

An Example table

	C	T	B
	DB	Zvi	Oracle
	DB	Zvi	Linux
	DB	Dennis	Oracle
	DB	Dennis	Linux
	OS	Dennis	Windows
	OS	Dennis	Linux
	OS	Jinyang	Windows
	OS	Jinyang	Linux

- This instance (and therefore the table in general) does not satisfy any functional dependencies
  - CT does not functionally determine B
  - CB does not functionally determine T
  - TB does not functionally determent C

Redundancies

	C	T	B
	DB	Zvi	Oracle
	DB	Zvi	Linux
	DB	Dennis	?
	DB	Dennis	?
	OS	Dennis	Windows
	OS	Dennis	Linux
	OS	Jinyang	?
	OS	Jinyang	?

	C	T	B
	DB	Zvi	Oracle
	DB	?	Linux
	DB	Dennis	Oracle
	DB	?	Linux
	OS	Dennis	Windows
	OS	?	Linux
	OS	Jinyang	Windows
	OS	?	Linux

- There are obvious redundancies
- In both cases, we know exactly how to fill the missing data if it was erased
- We decompose to get rid of anomalies

Decomposition

	C	T	B
	DB	Zvi	Oracle
	DB	Zvi	Linux
	DB	Dennis	Oracle
	DB	Dennis	Linux
	OS	Dennis	Windows
	OS	Dennis	Linux
	OS	Jinyang	Windows
	OS	Jinyang	Linux

	C	T
	DB	Zvi
	DB	Dennis
	OS	Dennis
	OS	Jinyang

	C	B
	DB	Oracle
	DB	Linux
	OS	Windows
	OS	Linux

Multivalued Dependencies And 4NF

- We had the following situation
- For each value of C there was
  - A set of values of T
  - A set of values of B
- Such that, every T of C had to appear with every B of C

  This is stated here rather loosely, but it is clear what it means
- The notation for this is: $C \twoheadrightarrow T \mid B$

- The tables CT and CB where in ***Fourth Normal Form (4NF)***

Now: To Algorithmic Techniques

- So far, our treatment was not algorithmic and we just looked at an interesting case exploring within the context of that case 3 issues

1. Avoiding (some) redundancies by converting tables to 3NF (and sometimes getting BCNF)
2. Preserving dependencies/constraints by making sure that dependencies (business rules) can be easily checked and enforced
3. Making sure that the decomposition of tables to obtain tables in better form does not cause us to lose information (lossless join) decomposition

- But we did not have an algorithmic procedure to do this
- We now continue with building up intuition and actually learning an algorithmic procedure

Closures Of Sets Of Attributes (Column Names)

- Closure of a set of attributes is an easy to use but extremely powerful tool for everything that follows
- "On the way" we may review some concepts
- We return to our old example, in which we are given a table with three columns (attributes)
  - Employee (E, for short, meaning really the SSN of the employee)
  - Grade (G, for short)
  - Salary (S, for short)
- Satisfies:
  1. $E \rightarrow G$
  2. $G \rightarrow S$
- We would like to find all the keys of this table
- A key is a minimal set of attributes, such that the values of these attributes, "force" some values for all the other attributes

Closures Of Sets Of Attributes

- In general, we have a concept of ***the closure of a set of attributes***
- Let X be a set of attributes, **then X^+ is the set of all attributes, whose values are forced by the values of X**
- In our example
  - E^+ = EGS (because given E we have the value of G and then because we have the value for G we have the value for S)
  - G^+ = GS
  - S^+ = S
- This is interesting because we have just showed that E is a key
- And here we could also figure out that this is the only key, as GS^+ = GS, so we will never get E unless we already have it
- Note that GS^+ really means $(GS)^+$ **and not** $G(S)^+$

Computing Closures Of Sets Of Attributes

- There is a very simple algorithm to compute X^+

  1. Let $Y = X$
  2. Whenever there is an FD, say $V \rightarrow W$, such that
     1. $V \subseteq Y$, and
     2. $W - Y$ is not empty

     add $W - Y$ to Y
  3. At termination $Y = X^+$

- The algorithm is very efficient
- Each time we look at all the functional dependencies
  - Either we can apply at least one functional dependency and make Y bigger (the biggest it can be are all attributes), or
  - We are finished

Example

- Let R = ABCDEGHIJK
- Given FDs:
  1. $K \rightarrow BG$
  2. $A \rightarrow DE$
  3. $H \rightarrow AI$
  4. $B \rightarrow D$
  5. $J \rightarrow IH$
  6. $C \rightarrow K$
  7. $I \rightarrow J$
- We will compute: ABC^+
  1. We start with $ABC^+ = ABC$
  2. Using FD number 2, we now have: $ABC^+ = ABCDE$
  3. Using FD number 6, we now have $ABC^+ = ABCDEK$
  4. Using FD number 1, we now have $ABC^+ = ABCDEKG$

  No FD can be applied productively anymore and we are done

Keys Of Tables

- ***The notion of an FD allows us to formally define keys***
- Given R (relation schema which is always denoted by its set of attributes), satisfying a set of FDs, a set of attributes X of R is a **key**, if and only if:
  - $X^+ = R$.
  - For any $Y \subseteq X$ such that $Y \neq X$, we have $Y^+ \neq R$.
- Note that if R does not satisfy any (nontrivial) FDs, then R is the only key of R
- "Trivial" means $P \rightarrow Q$ and $Q \subseteq P$: we saying something that is always true and not interesting
- Example, $AB \rightarrow A$ is always true and does not say anything interesting

- Example, if a table is R(FirstName,LastName) without any functional dependencies, then its key is just the pair (FirstName,LastName)

Keys of Tables

- If we apply our algorithm to the EGS example given earlier, we can now just compute that E was (the only) key by checking all the subsets of {E,G,S}

- Of course, in general, our algorithm is not efficient, but in practice what we do will be very efficient (most of the times)

Example: Airline Scheduling

- We have a table PFDT, where
  - **P**ILOT
  - **F**LIGHT NUMBER
  - **D**ATE
  - SCHEDULED_**T**IME_of_DEPARTURE

- The table satisfies the FDs:
  - $F \rightarrow T$
  - $PDT \rightarrow F$
  - $FD \rightarrow P$

Computing Keys

- We will compute all the keys of the table
- In general, this will be an exponential-time algorithm in the size of the problem
- But there will be useful heuristic making this problem tractable in practice
- We will introduce some heuristics here and additional ones later

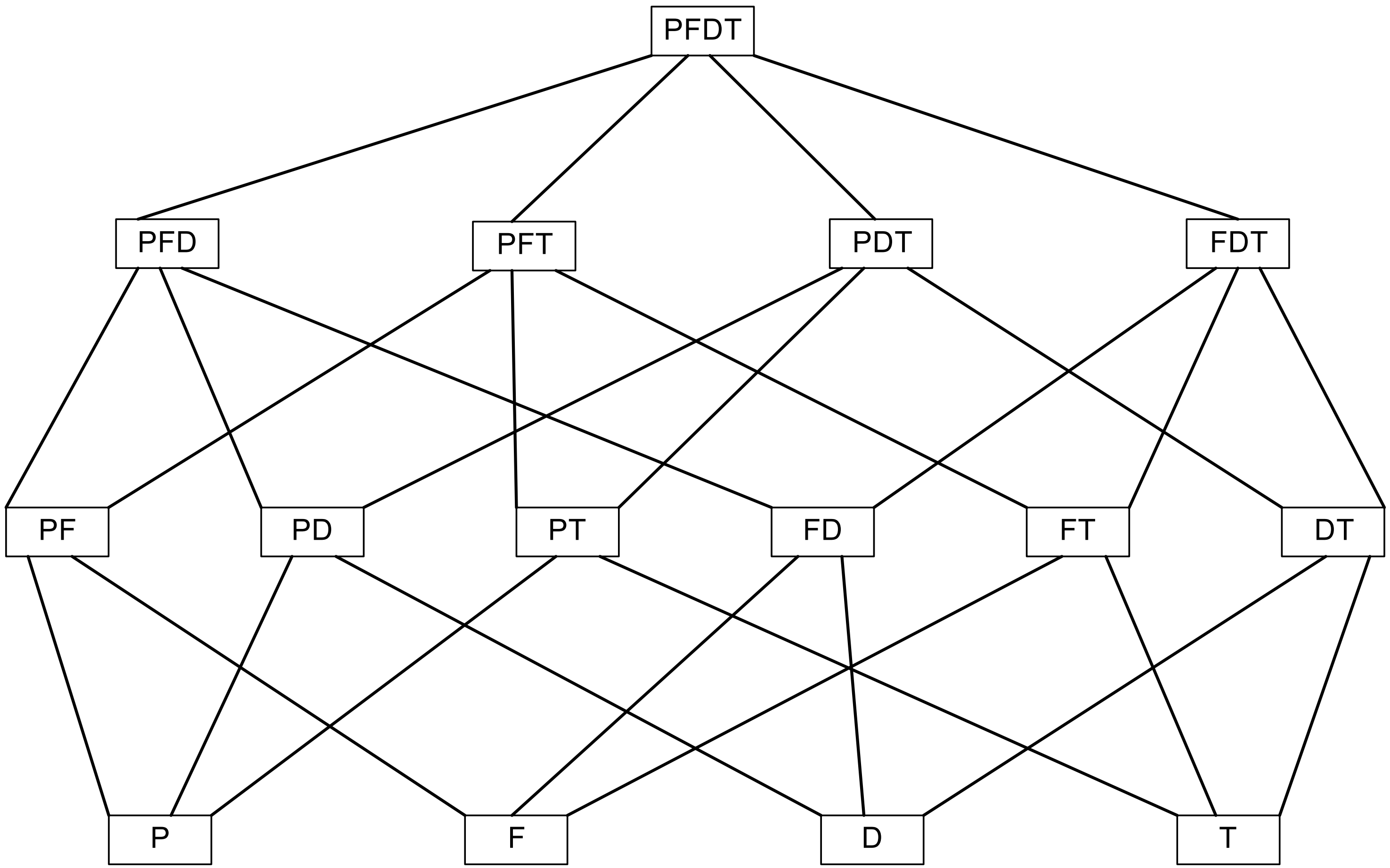
- We note that if some subset of attributes is a key, then no proper superset of it can be a key as it would not be minimal and would have superfluous attributes

Lattice Of Sets Of Attributes

- There is a natural structure (technically a lattice) to all the nonempty subsets of attributes
- I will draw the lattice here, in practice this is not done
  - Not necessary and too big
- We will look at all the non-empty subsets of attributes
- There are 15 of them: $2^4 - 1$

- The structure is clear from the drawing

Lattice Of Nonempty Subsets



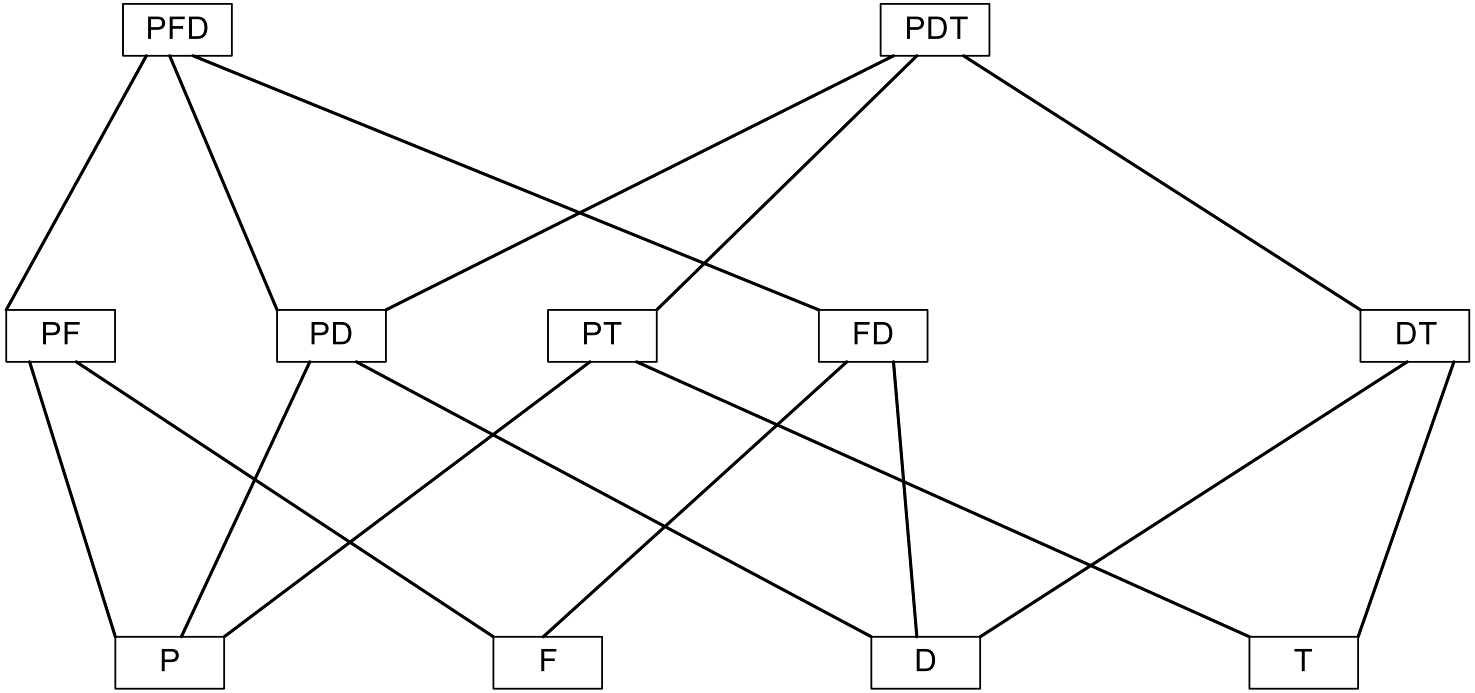
Keys Of PFDT

- The algorithm proceeds from bottom up
- We first try all potential 1-attribute keys, by examining all 1-attribute sets of attributes
  - $P^+ = P$
  - $F^+ = FT$
  - $D^+ = D$
  - $T^+ = T$

  There are no 1-attribute keys

- Note, that it is impossible for a key to have ***both*** F and T
  - Because if F is in a key, T will be automatically determined as it is included in the closure of F
- Therefore, we can prune our lattice

Pruned Lattice



Keys Of PFDT

- We try all potential 2-attribute keys
  - $PF^+ = PFT$
  - $PD^+ = PD$
  - $PT^+ = PT$
  - FD+ = FDPT
  - DT+ = DT

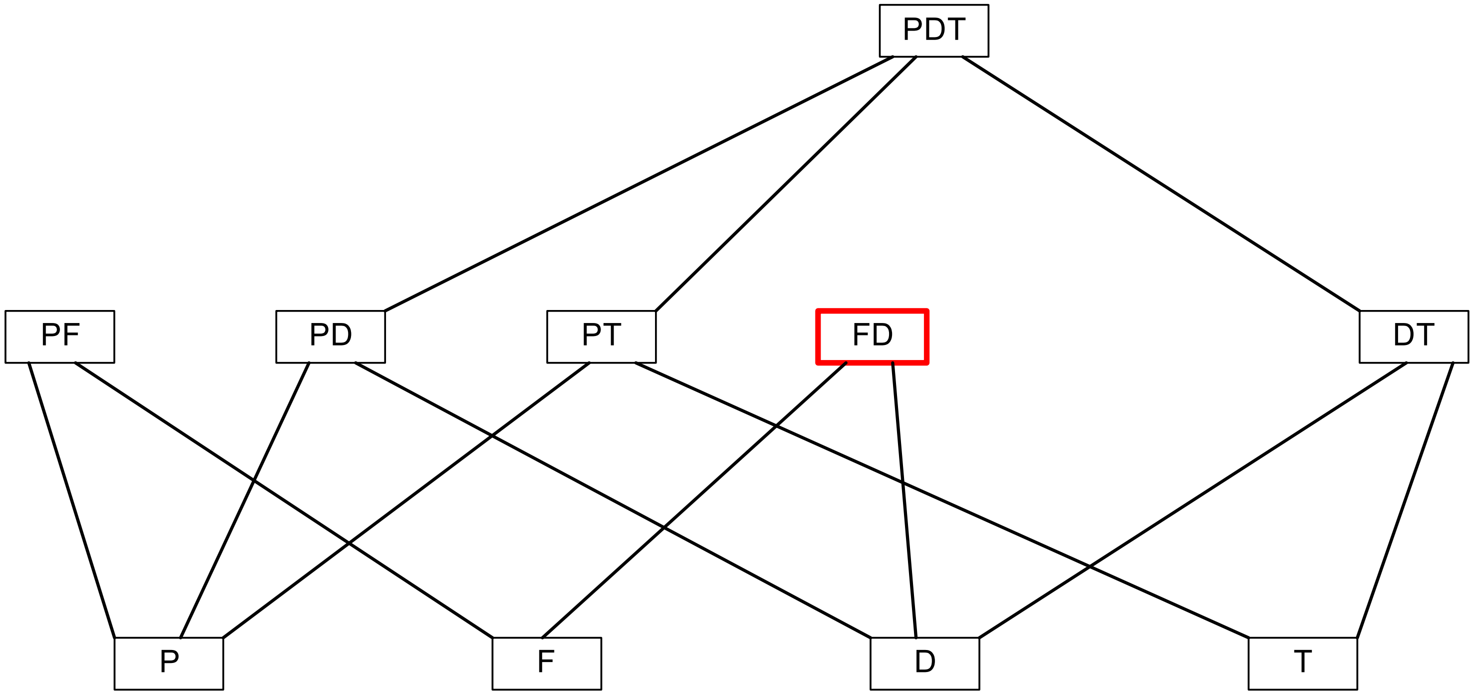
  There is one 2-attribute key: FD

- We can mark the lattice

- We can prune the lattice
  - Remove everything “above” FD because FD is key

Marked And Pruned Lattice

- The key we found is marked with red
- Some nodes can be removed



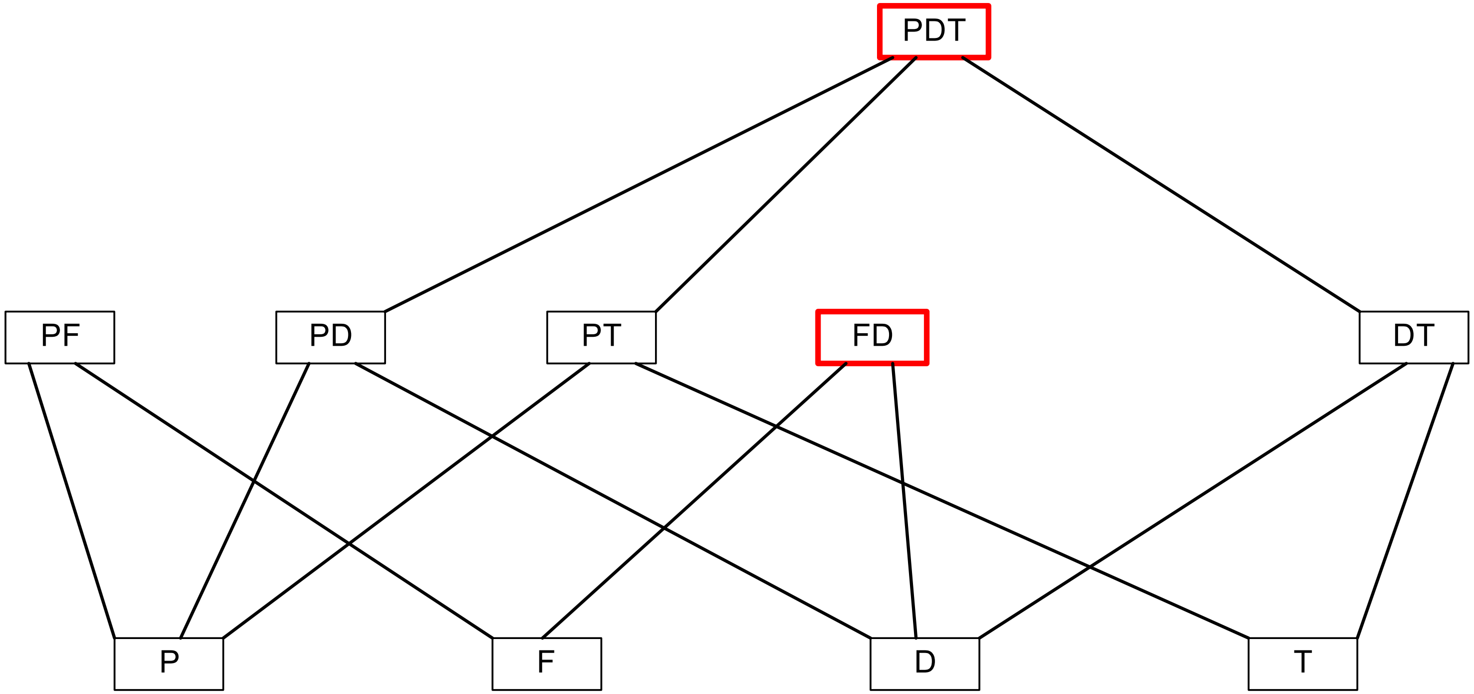
Keys Of PFDT

- We try all potential 3-attribute keys
  - $PDT^+ = PDTF$

  There is one 3-attribute key: PDT

Final Lattice
We Only Care About The Keys

- We could have removed some nodes, but we did not need to do that as we found all the possible keys



Finding A Decomposition

- Next, we will discuss by means of an example how to decompose a table into tables, such that

1. The decomposition is lossless join
2. Dependencies are preserved
3. Each resulting table is in 3NF

- Although this will be an example, the example will be sufficiently general so that the general procedure will be covered

The EmToPrHoSkLoRo Table

- The table deals with employees who use tools on projects and work a certain number of hours per week
- An employee may work in various locations and has a variety of skills
- All employees having a certain skill and working in a certain location meet in a specified room once a week

- The attributes of the table are:
  - Em: Employee
  - To: Tool
  - Pr: Project
  - Ho: Hours per week
  - Sk: Skill
  - Lo: Location
  - Ro: Room for meeting

The FDs Of The Table

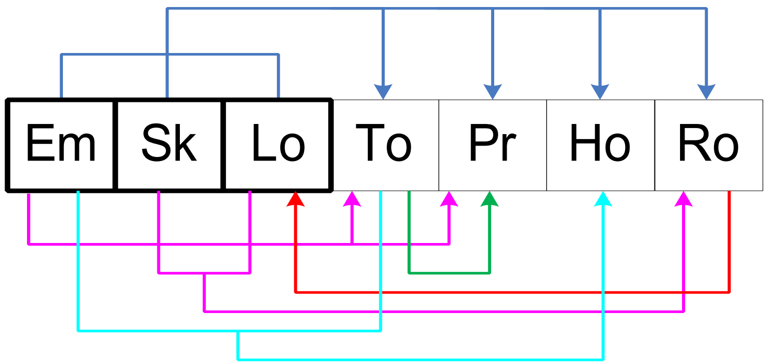
- The table deals with employees who use tools on projects and work a certain number of hours per week
- An employee may work in various locations and has a variety of skills
- All employees having a certain skill and working in a certain location meet in a specified room once a week
- The table satisfies the following FDs:
  - Each employee uses a single tool: Em → To
  - Each employee works on a single project: Em → Pr
  - Each tool can be used on a single project only: To → Pr
  - An employee uses each tool for the same number of hours each week: EmTo → Ho
  - All the employees working in a location having a certain skill always work in the same room (in that location): SkLo → Ro
  - Each room is in one location only: Ro → Lo

Sample Instance: Many Redundancies

	Em	To	Pr	Ho	Sk	Lo	Ro
	Mary	Pen	Research	20	Clerk	Boston	101
	Mary	Pen	Research	20	Writer	Boston	102
	Mary	Pen	Research	20	Writer	Buffalo	103
	Fang	Pen	Research	30	Clerk	New York	104
	Fang	Pen	Research	30	Editor	New York	105
	Fang	Pen	Research	30	Economist	New York	106
	Fang	Pen	Research	30	Economist	Buffalo	107
	Lakshmi	Oracle	Database	40	Analyst	Boston	101
	Lakshmi	Oracle	Database	40	Analyst	Buffalo	108
	Lakshmi	Oracle	Database	40	Clerk	Buffalo	107
	Lakshmi	Oracle	Database	40	Clerk	Boston	101
	Lakshmi	Oracle	Database	40	Clerk	Albany	109
	Lakshmi	Oracle	Database	40	Clerk	Trenton	110
	Lakshmi	Oracle	Database	40	Economist	Buffalo	107

Our FDs

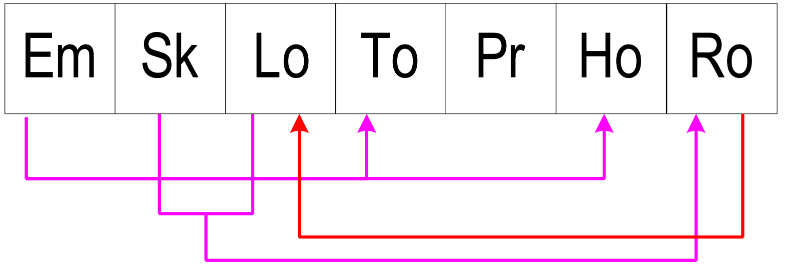
1. Em → To
2. Em → Pr
3. To → Pr
4. EmTo → Ho
5. SkLo → Ro
6. Ro → Lo



- What should we do with this drawing? I do not know. We need an algorithm
- We know how to find keys (we will actually do it later) and we can figure that EmSkLo could serve as the primary key, so we could draw using the appropriate colors
- But note that there for FD number 4, the left hand side contains an attribute from the key and an attribute from outside the key, so I used a new color
- Let's forget for now that I have told you what the primary key was, we will find it later

1: Getting A Minimal Cover

- We need to "simplify" our set of FDs to bring it into a "nicer" form, so called ***minimal cover*** or (sometimes called also ***canonical cover***)
- But, of course, the power has to be the same as we need to enforce the same business rules
- The algorithm for this will be covered later, it is very important
- The end result is:
  1. Em → ToHo
  2. To → Pr
  3. SkLo → Ro
  4. Ro → Lo
- From these we will build our tables directly, but just for fun, we can look at a drawing



2: Creating Tables From a Minimal Cover

- Create a table for each functional dependency
- We obtain the tables:
  1. EmToHo
  2. ToPr
  3. SkLoRo
  4. LoRo

3: Removing Redundant Tables

- LoRo is a subset of SkLoRo, so we remove it
- We obtain the tables:
  1. EmToHo
  2. ToPr
  3. SkLoRo

4: Ensuring The Storage Of The Global Key (Of The Original Table)

- We need to have a table containing the global key
- Perhaps one of our tables contains such a key
- So we check if any of them already contains a key of EmToPrHoSkLoRo:

1. EmToHo — $EmToHo^+$ = EmToHoPr, does not contain a key
2. ToPr — $ToPr^+$ = ToPr, does not contain a key
3. SkLoRo — $SkLoRo^+$ = SkLoRo, does not contain a key

- We need to add a table whose attributes form a global key

Finding Keys Using a Good Heuristic

- Let us list the FDs again (or could have worked with the minimal cover, does not matter):
  - Em → To
  - Em → Pr
  - To → Pr
  - EmTo → Ho
  - SkLo → Ro
  - Ro → Lo
- We can classify the attributes into 4 classes:
  1. Appearing on both sides of FDs; here To, Lo, Ro.
  2. Appearing on left sides only; here Em, Sk.
  3. Appearing on right sides only; here Pr, Ho.
  4. Not appearing in FDs; here none.

Finding Keys

- Facts:
  - Attributes of class 2 and 4 must appear in every key
  - Attributes of class 3 do not appear in any key
  - Attributes of class 1 may or may not appear in keys
- An algorithm for finding keys relies on these facts
  - Unfortunately, in the worst case, exponential in the number of attributes
- Start with the attributes in classes 2 and 4, add as needed (going bottom up) attributes in class 1, and ignore attributes in class 3

Finding Keys

- In our example, therefore, every key must contain EmSk
- To see, which attributes, if any have to be added, we compute which attributes are determined by EmSk
- We obtain
  - $EmSk^+$ = EmToPrHoSk
- Therefore Lo and Ro are missing
- It is easy to see that the table has two keys
  - EmSkLo
  - EmSkRo

Finding Keys

- Although not required strictly by the algorithm (which does not mind decomposing a table in 3NF into tables in 3NF) we can check if the original table was in 3NF
- We conclude that the original table is not in 3NF, as for instance, To $\rightarrow$ Pr is a transitive dependency and therefore not permitted for 3NF

4: Ensuring The Storage Of The Global Key

- None of the tables contains either EmSkLo or EmSkRo.
- Therefore, one more table needs to be added. We have 2 choices for the final decomposition
  1. EmToHo; satisfying Em → ToHo; primary key: Em
  2. ToPr; satisfying To → Pr; primary key To
  3. SkLoRo; satisfying SkLo → Ro and Ro → Lo; primary key SkLo or SkRo
  4. EmSkLo; not satisfying anything; primary key EmSkLo

  or

  1. EmToHo; satisfying Em → ToHo; primary key: Em
  2. ToPr; satisfying To → Pr; primary key To
  3. SkLoRo; satisfying SkLo → Ro and Ro → Lo; primary key SkLo or SkRo
  4. EmSkRo ; not satisfying anything; primary key EmSkRo
- We have completed our process and got a decomposition with the properties we needed; actually more than one

Computing Minimal Cover

- What remains to be done is to learn how to start with a set of FDs and to "reduce" them to a "clean" set ***with equivalent constraints power***
- This "clean" set is a minimal cover
- So we need to learn how to do that next

- We need first to understand better some properties of FDs

Relative Power Of Some FDs
H → G vs. H → GE

- Let us look at an example first
- Consider some table talking about employees in which there are three columns:
  1. Grade
  2. Bonus
  3. Salary
- Consider now two possible FDs (functional dependencies)
  1. Grade → Bonus
  2. Grade → Bonus Salary
- FD (2) is more restrictive, fewer relations will satisfy FD (2) than satisfy FD (1)
  - So FD (2) is stronger
  - Every relation that satisfies FD (2), must satisfy FD (1)
  - And we know this just because {Bonus} is a proper subset of {Bonus, Salary}

Relative Power Of Some FDs *$H \rightarrow G$ vs. $H \rightarrow GE$*

- An important note: $H \rightarrow GE$ is always at least as powerful as $H \rightarrow G$

that is

- If a relation satisfies $H \rightarrow GE$ it must satisfy $H \rightarrow G$

- What we are really saying is that if $GE = f(H)$, then of course $G = f(H)$
- An informal way of saying this: if being equal on H forces to be equal on GE, then of course there is equality just on G

- More generally, if X, Y, Z, are sets of attributes and $Z \subseteq Y$; then if $X \rightarrow Y$ is true then $X \rightarrow Z$ is true

Relative Power Of Some FDs
$A \rightarrow C$ vs. $AB \rightarrow C$

- Let us look at another example first
- Consider some table talking about employees in which there are three columns:
  1. Grade
  2. Location
  3. Salary
- Consider now two possible FDs
  1. Grade → Salary
  2. Grade Location → Salary
- FD (2) is less restrictive, more relations will satisfy FD (2) than satisfy FD (1)
  - So FD (1) is stronger
  - Every relation that satisfies FD (1), must satisfy FD (2)
  - And we know this just because {Grade} is a proper subset of {Grade, Salary}

Relative Power Of Some FDs
$A \rightarrow C$ vs. $AB \rightarrow C$

- An important note: $A \rightarrow C$ is always at least as powerful as $AB \rightarrow C$

that is

- If a relation satisfies $A \rightarrow C$ it must satisfy $AB \rightarrow C$

- What we are really saying is that if $C = f(A)$, then of course $C = f(A,B)$
- An informal way of saying this: if just being equal on A forces to be equal on C, then if we ***in addition*** know that there is equality on B also, of course it is still true that there is equality on C

- More generally, if X, Y, Z, are sets of attributes and $X \subseteq Y$; then if $X \rightarrow Z$ is true than $Y \rightarrow Z$ is true

Decomposition and Union of some FDs

- An FD $X \rightarrow A_1 A_2 \ldots A_m$, where A_i's are individual attributes

  is equivalent to

  the set of FDs:
  $X \rightarrow A_1$
  $X \rightarrow A_2$
  ...,
  $X \rightarrow A_m$

- Example
  FirstName LastName → Address Salary
  is equivalent to the set of the two FDs:
  Firstname LastName → Address
  Firstname LastName → Salary

Logical implications of FDs

- It will be important to us to determine if a given set of FDs ***forces*** some other FDs to be true
- Consider again the EGS relation
- Which FDs are satisfied?
  - $E \rightarrow G$, $G \rightarrow S$, $E \rightarrow S$ are all true in the real world
- If the real world tells you only:
  - $E \rightarrow G$ and $G \rightarrow S$
- Can you deduce on your own (and is it even always true?), ***without understanding the semantics of the application***, that
  - $E \rightarrow S$?

Logical implications of FDs

- Yes, by simple logical argument: transitivity
  1. Take any (set of) tuples that are equal on E
  2. Then given $E \rightarrow G$ we know that they are equal on G
  3. Then given $G \rightarrow S$ we know that they are equal on S
  4. So we have shown that $E \rightarrow S$ must hold

- We say that $E \rightarrow G$, $G \rightarrow S$ ***logically imply*** $E \rightarrow S$ and we write
- $E \rightarrow G, \ G \rightarrow S \ \models \ E \rightarrow S$

- This means:

  If a relation satisfies $E \rightarrow G$ and $G \rightarrow S$,

  then

  It must satisfy $E \rightarrow S$

Logical implications of FDs

- If the real world tells you only:
  - $E \rightarrow G$ and $E \rightarrow S$,
- Can you deduce on your own, without understanding the application that
  - $G \rightarrow S$
- No, because of a counterexample:

EGS	<u>E</u>	G	S
	Alpha	A	1
	Beta	A	2

- This relation satisfies $E \rightarrow G$ and $E \rightarrow S$, but violates $G \rightarrow S$
- For intuitive explanation, think: G means Height and S means Weight

Conclusion/Question

- Consider a relation EGS for which the three constraints E $\rightarrow$ G, G $\rightarrow$ S, and E $\rightarrow$ S ***must all be obeyed***

- ***It is enough*** to make sure that the two constraints E $\rightarrow$ G and G $\rightarrow$ S are not violated

- ***It is not enough*** to make sure that the two constraints E $\rightarrow$ G and E $\rightarrow$ S are not violated

- But what to do in general, large, complex cases?

To Remind: Closures Of Sets Of Attributes

- We consider some relation schema, which is a set of attributes, R (say EGS, which could also write as R(EGS))
- A set F of FDs for this schema (say $E \rightarrow G$ and $G \rightarrow S$)
- We take some $X \subseteq R$ (Say just the attribute E)
- We ask if two tuples are equal on X, what is the largest set of attributes on which they must be equal
- We call this set ***the closure of* X *with respect to* F** and denote it by $\mathbf{X_F^+}$ (in our case $E_F^+ = EGS$ and $S_F^+ = S$, as is easily seen)
- If it is understood what F is, we can write just $\mathbf{X^+}$

Towards A Minimal Cover

- This form will be based on trying to store a "concise" representation of FDs
- We will try to find a "small" number of "small" relation schemas that are sufficient to maintain the FDs
- The core of this will be to find "concise" description of FDs
  - Example: in ESG, E $\rightarrow$ S was not needed
- We will compute a ***minimal cover*** for a set of FDs
- The basic idea, simplification of a set of FDs by
  - Combining FDs when possible
  - Getting rid of unnecessary attributes
- We will start with examples to introduce the concepts and the tools

Union Rule: Combining Right Hand Sides (RHSs)

- $F = \{ AB \rightarrow C, AB \rightarrow D \}$

  is equivalent to

  $H = \{ AB \rightarrow CD \}$

- We have discussed this rule before
- Intuitively clear
- Formally we need to prove 2 things
  - F |= H is true; we do this (as we know) by showing that AB_F^+ contains CD; easy exercise
  - H |= F is true; we do this (as we know) by showing that AB_H^+ contains C and AB_H^+ contains D; easy exercise

- Note: you ***cannot*** combine LHSs based on equality of RHS and get an equivalent set of FDS
  - $F = \{A \rightarrow C, B \rightarrow C\}$ ***is stronger than*** $H = \{AB \rightarrow C\}$

Union Rule: Combining Right Hand Sides (RHSs)

- Stated formally:
  F = { X → Y, X → Z } ***is as powerful as*** H = { X → YZ }

- Easy proof, we omit

Relative Power Of FDs: Left Hand Side (LHS)

- $F = \{ AB \rightarrow C \}$

  is weaker than

  $H = \{ A \rightarrow C \}$

- We have discussed this rule before when we started talking about FDs
- Intuitively clear: in F, if we assume more (equality on both A and B) to conclude something (equality on C) then our FD is applicable in fewer case (does not work if we have equality is true on B's but not on C'S) and therefore F is weaker than H
- Formally we need to prove two things
  - F |= H is false; we do this (as we know) by showing that A_F^+ does not contain C; easy exercise
  - H |= F is true; we do this (as we know) by showing that AB_H^+ contains C; easy exercise

Relative Power Of FDs: Left Hand Side (LHS)

- Stated formally:
  $F = \{ XB \rightarrow Y \}$ ***is weaker than*** $H = \{ X \rightarrow Y \}$, (if $B \notin X$)

- Easy proof, we omit

- Can state more generally, replacing B by a set of attributes, but we do not need this

Relative Power Of FDs: Right Hand Side (RHS)

- $F = \{ A \rightarrow BC \}$

  is stronger than

  $H = \{ A \rightarrow B \}$

- Intuitively clear: in H, we deduce less from the same assumption, equality on A's
- Formally we need to prove two things
  - F |= H is true; we do this (as we know) by showing that A_F^+ contains B; easy exercise
  - H |= F is false; we do this (as we know) by showing that A_H^+ does not contain C; easy exercise

Relative Power Of FDs: Right Hand Side (RHS)

- Stated formally:
  $F = \{ X \rightarrow YC \}$ ***is stronger than*** $H = \{ X \rightarrow Y \}$, (if $C \notin Y$ and $C \notin X$)

- Easy proof, we omit

- Can state more generally, replacing C by a set of attributes, but we do not need this

Simplifying Sets Of FDs

- At various stages of the algorithm we will have
  - An “old” set of FDs
  - A “new” set of FDs
- The two sets will not vary by “very much”
- We will indicate the parts that do not change by . . .
- Of course, as we are dealing with sets, the order of the FDs in the set does not matter

Simplifying Set Of FDs
By Using The Union Rule

- X, Y, Z are sets of attributes
- Let F be:

  …
  $X \rightarrow Y$
  $X \rightarrow Z$

- Then, F is equivalent to the following H:

  …
  $X \rightarrow YZ$

Simplify Set Of FDS By Simplifying LHS

- Let X, Y are sets of attributes and B a single attribute not in X
- Let F be:

  …
  $XB \rightarrow Y$

- Let H be:

  …
  $X \rightarrow Y$

- Then if $F \models X \rightarrow Y$ holds, then we can replace F by H without changing the "power" of F
- We do this by showing that X_F^+ contains Y
  - H could only be stronger, but we are proving it is not actually stronger, but equivalent

Simplify Set Of FDS By Simplifying LHS

- H can only be stronger than F, as we have replaced a weaker FD by a stronger FD
- But if we F |= H holds, this “local” change does not change the overall power
- Example below
- Replace
  - $AB \rightarrow C$
  - $A \rightarrow B$

    by
  - $A \rightarrow C$
  - $A \rightarrow B$

Simplify Set Of FDS By Simplifying RHS

- Let X, Y are sets of attributes and C a single attribute not in Y
- Let F be:

  …
  $X \rightarrow YC$
  …

- Let H be:

  …
  $X \rightarrow Y$
  …

- Then if H |= $X \rightarrow YC$ holds, then we can replace F by H without changing the "power" of F
- We do this by showing that X_H^+ contains YC
  - H could only be weaker, but we are proving it is not actually weaker, but equivalent

Simplify Set Of FDS By Simplifying RHS

- H can only be weaker than F, as we have replaced a stronger FD by a weaker FD
- But if H |= F holds, this “local” change does not change the overall power
- Example below
- Replace
  - $A \rightarrow BC$
  - $B \rightarrow C$

  by
  - $A \rightarrow B$
  - $B \rightarrow C$

Minimal Cover

- Given a set of FDs F, find a set of FDs F_m, that is (in a sense we formally define later) minimal
- Algorithm:

1. Start with F
2. Remove all trivial functional dependencies
3. Repeatedly apply (in whatever order you like), until no changes are possible
   - Union Simplification (it is better to do it as soon as possible, whenever possible)
   - RHS Simplification
   - LHS Simplification
4. What you get is a minimal cover

- We proceed through a largish example to exercise all possibilities

The EmToPrHoSkLoRo Relation

- The relation deals with employees who use tools on projects and work a certain number of hours per week
- An employee may work in various locations and has a variety of skills
- All employees having a certain skill and working in a certain location meet in a specified room once a week

- The attributes of the relation are:
  - Em: Employee
  - To: Tool
  - Pr: Project
  - Ho: Hours per week
  - Sk: Skill
  - Lo: Location
  - Ro: Room for meeting

The FDs Of The Relation

- The relation deals with employees who use tools on projects and work a certain number of hours per week
- An employee may work in various locations and has a variety of skills
- All employees having a certain skill and working in a certain location meet in a specified room once a week
- The relation satisfies the following FDs:
  - Each employee uses a single tool: Em → To
  - Each employee works on a single project: Em → Pr
  - Each tool can be used on a single project only: To → Pr
  - An employee uses each tool for the same number of hours each week: EmTo → Ho
  - All the employees working in a location having a certain skill always work in the same room (in that location): SkLo → Ro
  - Each room is in one location only: Ro → Lo

Sample Instance

	Em	To	Pr	Ho	Sk	Lo	Ro
	Mary	Pen	Research	20	Clerk	Boston	101
	Mary	Pen	Research	20	Writer	Boston	102
	Mary	Pen	Research	20	Writer	Buffalo	103
	Fang	Pen	Research	30	Clerk	New York	104
	Fang	Pen	Research	30	Editor	New York	105
	Fang	Pen	Research	30	Economist	New York	106
	Fang	Pen	Research	30	Economist	Buffalo	107
	Lakshmi	Oracle	Database	40	Analyst	Boston	101
	Lakshmi	Oracle	Database	40	Analyst	Buffalo	108
	Lakshmi	Oracle	Database	40	Clerk	Buffalo	107
	Lakshmi	Oracle	Database	40	Clerk	Boston	101
	Lakshmi	Oracle	Database	40	Clerk	Albany	109
	Lakshmi	Oracle	Database	40	Clerk	Trenton	110
	Lakshmi	Oracle	Database	40	Economist	Buffalo	107

Our FDs

1. Em → To
2. Em → Pr
3. To → Pr
4. EmTo → Ho
5. SkLo → Ro
6. Ro → Lo

Run The Algorithm

- Using the union rule, we combine RHS of 1 and 2, getting:
  1. Em $\rightarrow$ ToPr
  2. To $\rightarrow$ Pr
  3. EmTo $\rightarrow$ Ho
  4. SkLo $\rightarrow$ Ro
  5. Ro $\rightarrow$ Lo

Run The Algorithm

- No RHS can be combined, so we check whether there are any redundant attributes.
- We start with FD 1, where we attempt to remove an attribute from RHS
  - We check whether we can remove To. This is possible if we can derive Em → To using

    Em → Pr

    To → Pr

    EmTo → Ho

    SkLo → Ro

    Ro → Lo

  Computing the closure of Em using the above FDs gives us only EmPr, so the attribute To must be kept.

Run The Algorithm

- We check whether we can remove Pr. This is possible if we can derive Em → Pr using

  Em → To

  To → Pr

  EmTo → Ho

  SkLo → Ro

  Ro → Lo

  Computing the closure of Em using the above FDs gives us EmToPrHo, so the attribute Pr is redundant

Run The Algorithm

- We now have
  1. Em → To
  2. To → Pr
  3. EmTo → Ho
  4. SkLo → Ro
  5. Ro → Lo
- No RHS can be combined, so we continue attempting to remove redundant attributes. The next one is FD 3, where we attempt to remove an attribute from LHS
  - We check if Em can be removed. This is possible if we can derive To → Ho using *all* the FDs. Computing the closure of To using the FDs gives ToPr, and therefore Em cannot be removed
  - We check if To can be removed. This is possible if we can derive Em → Ho using *all* the FDs. Computing the closure of Em using the FDs gives EmToPrHo, and therefore To can be removed

Run The Algorithm

- We now have
  1. Em → To
  2. To → Pr
  3. Em → Ho
  4. SkLo → Ro
  5. Ro → Lo
- We can now combine RHS of 1 and 3 and get
  1. Em → ToHo
  2. To → Pr
  3. SkLo → Ro
  4. Ro → Lo

Run The Algorithm

- We now attempt to remove an attribute from the LHS of 3, and an attribute from RHS of 1, but neither is possible
- Therefore we are done
- We have computed a minimal cover for the original set of FDs

Minimal Cover

- A set of FDs, F_m, is a minimal cover for a set of FD F,

  if and only if

1. F_m is minimal, that is
   1. No two FDs in it can be combined using the union rule
   2. No attribute can be removed from a RHS of any FD in F_m without changing the power of F_m
   3. No attribute can be removed from a LHS of any FD in F_m without changing the power of F_m
2. F_m is equivalent in power to F

- Note that there could be more than one minimal cover for F, as we have not specified the order of applying the simplification operations

How About EGS

- Applying the algorithm to EGS with
  1. E → G
  2. G → S
  3. E → S
- Using the union rule, we combine 1 and 3 and get
  1. E → GS
  2. G → S
- Simplifying RHS of 1 (this is the only attribute we can remove), we get
  1. E → G
  2. G → S
- We automatically got the two “important” FDs!

An Algorithm For "An Almost" 3NF Lossless-Join Decomposition

- Input: relation schema R and a set of FDs F
- Output: almost-decomposition of R into R_1, R_2, …, R_n, each in 3NF
- Algorithm

1. Produce F_m, a minimal cover for F
2. For each $X \rightarrow Y$ in F_m create a new relation schema XY
3. For every new relation schema that is a subset (including being equal) of another new relation schema (that is the set of attributes is a subset of attributes of another schema or the two sets of attributes are equal) remove this relation schema (the "smaller" one or one of the equal ones); but if the two are equal, need to keep one of them
4. The set of the remaining relation schemas is an "almost final decomposition"

Back To Our Example

- For our EmToPrHoSkLoRo example, we previously computed the following minimal cover:
  1. Em $\rightarrow$ ToHo
  2. To $\rightarrow$ Pr
  3. SkLo $\rightarrow$ Ro
  4. Ro $\rightarrow$ Lo

Creating Relations

- Create a relation for each functional dependency
- We obtain the relations:
  1. EmToHo
  2. ToPr
  3. SkLoRo
  4. LoRo

Removing Redundant Relations

- LoRo is a subset of SkLoRo, so we remove it
- We obtain the relations:
  1. EmToHo
  2. ToPr
  3. SkLoRo

Assuring Storage Of A Global Key

- If no relation contains a key of the original relation, add a relation whose attributes form such a key
- Why do we need to do this?
  - Because otherwise we may not have a decomposition
  - Because otherwise the decomposition may not be lossless

Why It Is Necessary To Store A Global Key Example

- Consider the relation LnFn:
  - Ln: Last Name
  - Fn: First Name
- There are no FDs
- The relation has only one key:
  - LnFn
- Our algorithm (without the key included) produces no relations
- A condition for a decomposition: Each attribute of R has to appear in at least one Ri
- So we did not have a decomposition
- But if we add the relation consisting of the attributes of the key
  - We get LnFn (this is fine, because the original relations had no problems and was in a good form, actually in BCNF, which is always true when there are no (nontrivial) FDs)

Assuring Storage Of A Global Key

- If no relation contains a key of the original relation, add a relation whose attributes form such a key
- It is easy to test if a "new" relation contains a key of the original relation
- Compute the closure of the relation with respect to all FDs (either original or minimal cover, it's the same) and see if you get all the attributes of the original relation
- If not, you need to find some key of the original relation
- We have studied this before

Returning to Our Example

- We pick the decomposition
  1. EmToHo
  2. ToPr
  3. SkLoRo
  4. EmSkLo

- We have the minimal set of FDs of the simplest form (before any combinations)
  1. Em $\rightarrow$ ToHo
  2. To $\rightarrow$ Pr
  3. SkLo $\rightarrow$ Ro
  4. Ro $\rightarrow$ Lo

Returning to Our Example

- Everything can be described as follows:
- The relations, their keys, and FDs that need to be explicitly mentioned are:
  1. EmToHo key: Em
  2. ToPr key: To
  3. SkLoRo key: SkLo, key SkRo, and functional dependency Ro → Lo
  4. EmSkLo key: EmSkLo
- In general, when you decompose as we did, a relation may have several keys and satisfy several FDs that do not follow from simply knowing keys
- In the example above there was one relation that had such an FD, which made it automatically not a BCNF relation (but by our construction a 3NF relation)

Back to SQL DDL

- How are we going to express in SQL what we have learned?
- We need to express:
  - keys
  - functional dependencies
- Expressing keys is very easy, we use the PRIMARY KEY and UNIQUE keywords
- Expressing functional dependencies is possible also by means of a CHECK condition
  - What we need to say for the relation SkLoRo is that each tuple satisfies the following condition

    There are no tuples in the relation with the same value of Ro and different values of Lo

Back to SQL DDL

- CREATE TABLE SkLoRo
```
(Sk …,
Lo …,
Ro…,
UNIQUE (Sk,Ro),
PRIMARY KEY (Sk,Lo),
CHECK (NOT EXISTS SELECT *
                FROM SkLoRo AS Copy
                WHERE (SkLoRo.Ro = Copy.Ro
                          AND NOT SkLoRo.Lo = Copy.Lo)));
```

- But this is generally not supported by actual relational database systems
- Even assertions are frequently not supported
- Can do it differently
- Whenever there is an insert or update, check that FDs hold, or reject these actions

Maintaining FDs During Insertion

- We have a table R satisfying some FDs
- We have a table T of “candidates” for inserting into R
- We want to construct a subset of U of T consisting only of those tuples whose insertion into R would not violate FDs

- We show how to do it for the simple example of R = EGS, where we need to maintain:
  - E is the primary key
  - G → S holds
- We replace

```
INSERT INTO R
(SELECT *
FROM T);
```

By the following

Maintaining FDs During Insertion

```
INSERT INTO R
  (SELECT *
  FROM T
  WHERE NOT EXISTS
    (SELECT *
     FROM R
     WHERE (R.G = T.G AND R.S <> T.S) OR (R.E = T.E)
    )
  );
```

- The WHERE condition will only insert those tuples from T to R that satisfy the conditions
  - There is no tuple in R with the same G but a different S
  - There is no tuple in R with the same value of the primary key E

Key Ideas

- Need for decomposition of tables
- Functional dependencies
- Some types of functional dependencies:
  - Partial dependencies
  - Transitive dependencies
  - Into full key dependencies
- First Normal Form: 1NF
- Second Normal Form: 2NF
- Third Normal Form: BCNF
- Removing redundancies
- Lossless join decomposition
- Preservation of dependencies
- 3NF vs. BCNF

Key Ideas

- Multivalued dependencies
- Fourth Normal Form: 4NF
- Minimal cover for a set of functional dependencies
- Algorithmic technique for finding keys
- Algorithmic technique for computing an a minimal cover
- Algorithmic technique for obtaining a decomposition of relation into a set of relations, such that
  - The decomposition is lossless join
  - Dependencies are preserved
  - Each resulting relation is in 3NF