

Functional Dependencies

政治大學

資訊科學系

沈錦坤

Outline

- ◆ Good Relation Schemas
- ◆ Informal Design Guidelines for Relational Schemas
- ◆ Functional Dependencies

Good Relation Schemas

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

WORKS_ON

Essn	Pno	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	NULL

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

Good Relation Schemas

- ◆ Theory to choose good relation schemas
 - To measure formally why one set of groupings of attributes into relation schemas is better than another
- ◆ 2 Levels of goodness
 - Logical level (conceptual level)
 - Enable users to understand clearly the meaning of the data in the relations & to formulate queries correctly
 - Implementation (storage level)
 - How the tuples in a base relation are stored & updated

Approaches of Database Design

- ◆ Approaches
 - Bottom up design (*requires experience*)
 - Consider the basic relationships among individual attributes as the starting point
 - Not very popular in practice
 - Suffers from the problem of collecting a large number of binary attribute relationships
 - Design by synthesis
 - Top-down design *
 - Start with a number of groupings of attributes into relations obtained from conceptual design
 - Further decomposition until all desirable properties are met
 - Design by analysis

Informal Design Guidelines for Relational Schemas

Informal Design Guidelines for Relation Schemas

- ◆ 4 Informal measures of quality for relation schema design
 - Semantics of the attributes
 - Reducing the redundant values in tuples
 - Reducing the null values in tuples
 - Disallowing the possibility of generating spurious tuples

Semantics of the Relation Attributes

- ◆ Clear semantics
- ◆ Guideline 1
 - Do not combine attributes from multiple entity types & relationship types into a single relation
 - If a relation schema corresponds to one entity type or one relationship type, the meaning tends to be clear

employee

ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER
-------	------------	-------	---------	---------

p.k.

department

DNAME	<u>DNUMBER</u>	DMGRSSN
-------	----------------	---------

p.k.

dept - locations

DNUMBER	DLOCATION
p.k.	

project

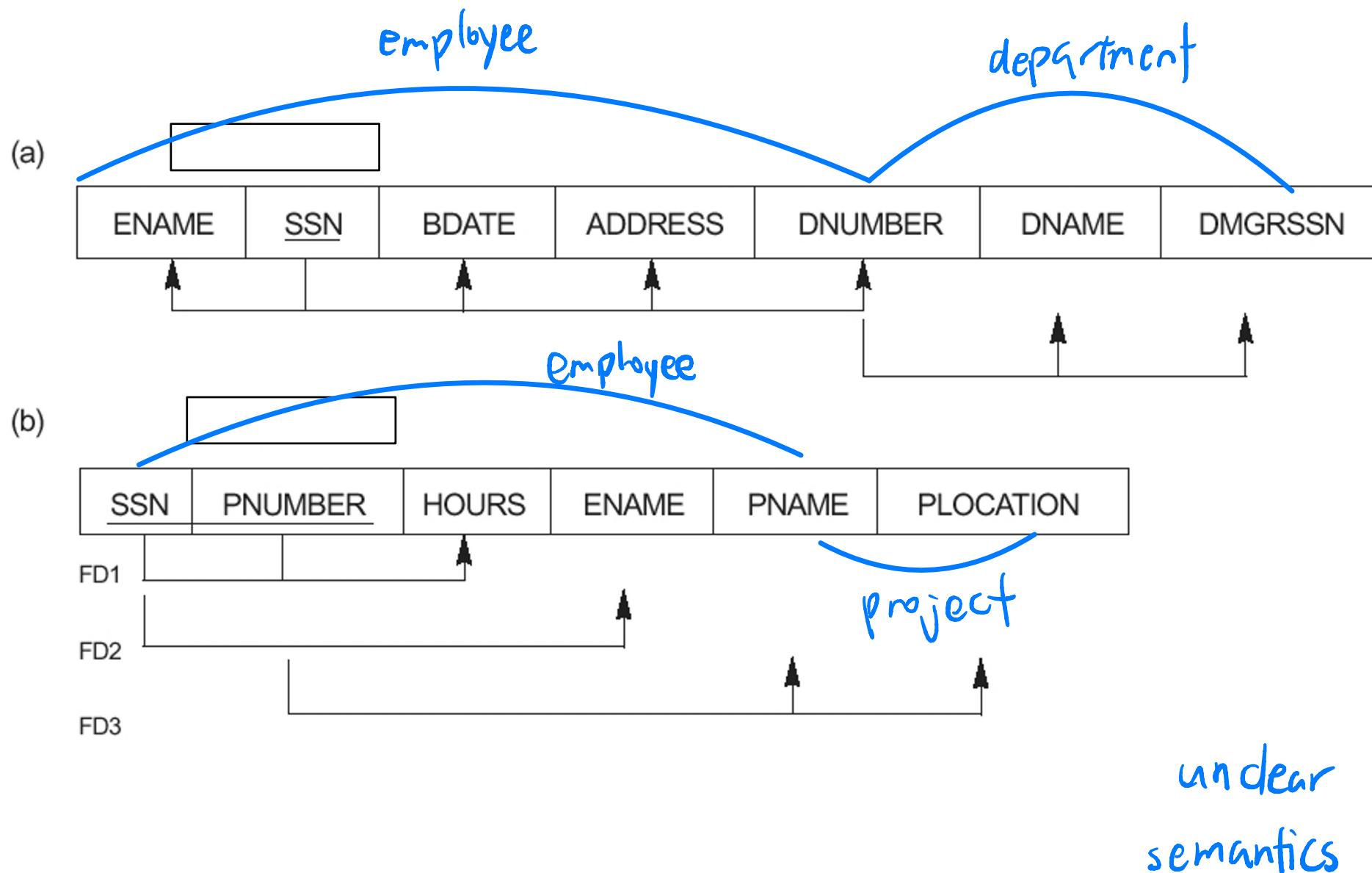
PNAME	<u>PNUMBER</u>	PLOCATION	DNUM
-------	----------------	-----------	------

p.k.

works - on

SSN	PNUMBER	HOURS
p.k.		

clear
Semantics



Redundant Information in Tuples & Update Anomalies

- ◆ Guideline 2
 - Design the base relation schemas so that no insertion, deletion or modification anomalies are present in the relations
 - If any anomalies are present, make sure that the programs that update the database will operate correctly

redundant
info.

EMP_DEPT

ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Research	333445555
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Research	333445555
Zelaya,Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Administration	987654321
Narayan,Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Research	333445555
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Administration	987654321
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1	Headquarters	888665555

- To insert a new department that has no employee yet ?
No (∴ pk B SSN)
- To delete an employee that happens to be the last employee working for a particular department ?
The department info. will be gone as well.
- To modify information of a department ?
It may produce unexpected result if some records are not updated in accident.

Redundant Information in Tuples & Update Anomalies (cont.)

- ◆ Grouping attributes into relation schemas
 - Storage space
 - Update anomalies
 - Insertion anomalies
 - Consistency of department information for insertion of employee
 - Difficult to insert a new department that has no employee yet
 - Deletion anomalies
 - Delete an employee that happens to be the last employee working for a particular department, department information is lost
 - Modification anomalies
 - Consistency of department information for modification of department information of one tuple

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

WORKS_ON

Essn	Pno	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	NULL

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

Null Values in Tuples

- ◆ Guideline 3
 - Avoiding placing attributes in a base relation whose values may frequently be null
 - If nulls are unavoidable, make sure that they apply in exceptional cases only & do not apply to a majority of tuples in the relation
- ◆ e.g.: if only 10% employees have individual offices, there is little justification for including an attribute Office_Number in the Employee relation

- ◆ e.g.

DEPARTMENT

Dname	<u>Dnumber</u>	Mgr_ssn	Mgr_start_date
-------	----------------	---------	----------------

EMPLOYEE

Fname	Minit	Lname	<u>Ssn</u>	Bdate	Address	Sex	Salary	Super_ssn	Dno
-------	-------	-------	------------	-------	---------	-----	--------	-----------	-----

Null Values in Tuples (cont.)

◆ Nulls

- When grouping attributes together into a fat relation, if many of the attributes do not apply to all tuples in the relation, there exist many nulls in those tuples
- Disadvantage
 - Waste storage space
 - Aggregation function
 - Different interpretation
 - **The attribute does not apply to this tuple**
 - **The attribute value for this tuple is unknown**
 - **The value is known but absent, it has not been recorded yet**

Generation of Spurious Tuples

- ◆ Spurious tuples
- ◆ Guidelines 4
 - Design relation schemas so that they can joined with equality conditions on attributes that are either primary keys or foreign keys in a way that guarantees that no spurious tuples are generated
 - Do not have relations that contain matching attributes other than foreign key-primary key combinations
 - If such relations are unavoidable, do not join them on such attributes.

EMP_LOCS

ENAME	PLOCATION
Smith, John B.	Bellaire
Smith, John B.	Sugarland
Narayan, Ramesh K.	Houston
English, Joyce A.	Bellaire
English, Joyce A.	Sugarland
Wong, Franklin T.	Sugarland
Wong, Franklin T.	Houston
Wong, Franklin T.	Stafford
Zelaya, Alicia J.	Stafford
Jabbar, Ahmad V.	Stafford
Wallace, Jennifer S.	Stafford
Wallace, Jennifer S.	Houston
Borg,James E.	Houston

EMP_LOCS . PLOCATION =
EMP_PROJ1 . PLOCATION

EMP_PROJ1

SSN	PNUMBER	HOURS	PNAME	PLOCATION
123456789	1	32.5	Product X	Bellaire
123456789	2	7.5	Product Y	Sugarland
666884444	3	40.0	Product Z	Houston
453453453	1	20.0	Product X	Bellaire
453453453	2	20.0	Product Y	Sugarland
333445555	2	10.0	Product Y	Sugarland
333445555	3	10.0	Product Z	Houston
333445555	10	10.0	Computerization	Stafford
333445555	20	10.0	Reorganization	Houston
999887777	30	30.0	Newbenefits	Stafford
999887777	10	10.0	Computerization	Stafford
987987987	10	35.0	Computerization	Stafford
987987987	30	5.0	Newbenefits	Stafford
987654321	30	20.0	Newbenefits	Stafford
987654321	20	15.0	Reorganization	Houston
888665555	20	null	Reorganization	Houston

does not make sense!

SSN	PNUMBER	HOURS	PNAME	PLOCATION	
123456789	1	32.5	ProductX	Bellaire	Smith,John B.
* 123456789	1	32.5	ProductX	Bellaire	English,Joyce A.
123456789	2	7.5	ProductY	Sugarland	Smith,John B.
* 123456789	2	7.5	ProductY	Sugarland	English,Joyce A.
* 123456789	2	7.5	ProductY	Sugarland	Wong,Franklin T.
666884444	3	40.0	ProductZ	Houston	Narayan,Ramesh K.
* 666884444	3	40.0	ProductZ	Houston	Wong,Franklin T.
* 453453453	1	20.0	ProductX	Bellaire	Smith,John B.
453453453	1	20.0	ProductX	Bellaire	English,Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Smith,John B.
453453453	2	20.0	ProductY	Sugarland	English,Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Wong,Franklin T.
* 333445555	2	10.0	ProductY	Sugarland	Smith,John B.
* 333445555	2	10.0	ProductY	Sugarland	English,Joyce A.
333445555	2	10.0	ProductY	Sugarland	Wong,Franklin T.
* 333445555	3	10.0	ProductZ	Houston	Narayan,Ramesh K.
333445555	3	10.0	ProductZ	Houston	Wong,Franklin T.
333445555	10	10.0	Computerization	Stafford	Wong,Franklin T.
* 333445555	20	10.0	Reorganization	Houston	Narayan,Ramesh K.
333445555	20	10.0	Reorganization	Houston	Wong,Franklin T.

•
•
•

Summary of Design Guidelines

- ◆ Anomalies imply additional work to be done during insertion into & modification of a relation
- ◆ Anomalies may cause accidental loss of information during a deletion from a relation
- ◆ Nulls wastes storage space and make difficulty of performing aggregation operations and joins due to null values
- ◆ Generation of invalid & spurious data during joins on improperly related base relations

Functional Dependencies

Functional Dependencies

- ◆ A functional dependency, $X \rightarrow Y$,
between two sets of attributes X & Y
that are subset of R
 - for any two tuples t_1 & t_2 in a relation state r of R
that have $t_1[X] = t_2[X]$,
they must also have $t_1[Y] = t_2[Y]$
 - The values of Y of a tuple in R depend on the value of X
 - The values of X component uniquely determine the values of Y component
 - Y is functional dependent on X

Functional Dependencies

- ◆ A functional dependency, $X \rightarrow Y$,
between two sets of attributes X & Y
 - If X is a candidate key of R
then $X \rightarrow Y$ for any subset of attributes Y of R
 - If $X \rightarrow Y$ in R , this does not say whether or not $Y \rightarrow X$ in R
 - FD cannot be inferred automatically
from a given relation extension R
but must be defined explicitly by someone
who knows the semantics of the attributes of R

EMP_DEPT

ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Research	333445555
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Research	333445555
Zelaya,Alicia J.	999887777	1968-07-19	3321 Castle, Spring,TX	4	Administration	987654321
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Administration	987654321
Narayan,Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Research	333445555
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Administration	987654321
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1	Headquarters	888665555

(a)

EMP_DEPT

ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN

```
graph TD; ENAME[ENAME] --> ENAMECell[ENAME]; SSN[SSN] --> SSNCCell[SSN]; BDATE[BDATE] --> BDATCell[BDATE]; ADDRESS[ADDRESS] --> ADDCell[ADDRESS]; DNUMBER[DNUMBER] --> DNumbCell[DNUMBER]; DNAME[DNAME] --> DNmeCell[DNAME]; DMGRSSN[DMGRSSN] --> DMGRSSNCell[DMGRSSN]
```

(b)

EMP_PROJ

SSN	PNUMBER	HOURS	ENAME	PNAME	PLOCATION
FD1					
FD2					
FD3					

```
graph TD; SSN[SSN] --> SSNCCell[SSN]; PNUMBER[PNUMBER] --> PNumbCell[PNUMBER]; HOURS[HOURS] --> HOURSCell[HOURS]; ENAME[ENAME] --> ENAMECell[ENAME]; PNAME[PNAME] --> PNmeCell[PNAME]; PLOCATION[PLOCATION] --> PLocCell[PLOCATION];
```

```
graph TD; FD1[FD1] --> PNUMBERCell[PNUMBER]; FD1 --> HOURSCell[HOURS]; FD2[FD2] --> ENAMECell[ENAME]; FD3[FD3] --> PNAMECell[PNAME]; FD3 --> PLocCell[PLOCATION]
```

Functional Dependencies (cont.)

- ◆ Trivial functional dependency
 - A functional dependency $X \rightarrow Y$
 - is trivial if $X \supseteq Y$
 - is nontrivial otherwise
 - e.g. $\{SSN, BDate\} \rightarrow \{BDate\}$ is a trivial FD
 - e.g. $\{SSN\} \rightarrow \{Bdate\}$ is a nontrivial FD
 - e.g. $\{Bdate\} \not\rightarrow \{SSN\}$ is not a FD

Inference Rules for Functional Dependencies

♦ Inference rules

- Reflexive rule: if $X \supseteq Y$, then $X \rightarrow Y$
- Augmentation rule: $\{X \rightarrow Y\} \models XZ \rightarrow YZ$
- Transitive rule: $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$
- Decomposition rule: $\{X \rightarrow YZ\} \models X \rightarrow Y$
- Union rule: $\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$
- Pseudotransitive rule: $\{X \rightarrow Y, WY \rightarrow Z\} \models WX \rightarrow Z$

Armstrong's
Rules

derived

Armstrong's Inference Rules

- ◆ Closure of F
 - The set of dependencies F^+
 - Can be determined from F by using only inference rules
IR1 thru IR3
 - IR1 thru IR3 are known as Armstrong's rules

Example of Closure of F

◆ $F =$

- $\{\text{SSN}\} \rightarrow \{\text{Ename}\}$,
- $\{\text{Pnumber}\} \rightarrow \{\text{Pname}, \text{PLocation}\}$
- $\{\text{SSN}, \text{PNumber}\} \rightarrow \{\text{Hours}\}$

◆ $F^+ =$

- $\{\text{SSN}\} \rightarrow \{\text{SSN}, \text{Ename}\}$ (*Augmentation Rule*)
- $\{\text{Pnumber}\} \rightarrow \{\text{Pnumber}, \text{Pname}, \text{Plocation}\}$ (*Augmentation Rule*)
- $\{\text{SSN}, \text{Pnumber}\} \rightarrow \{\text{SSN}, \text{Pnumber}, \text{Ename}, \text{Pname}, \text{Plocation}, \text{Hours}\}$

① $\{ \text{ssN} \} \rightarrow \{ \text{Ename} \}$ $\xrightarrow[\text{Rule}]{\text{Augmentation}} \{ \text{ssN} \} \rightarrow \{ \text{ssN}, \text{Ename} \}$

② $\{ \text{Pnumber} \} \rightarrow \{ \text{Pname}, \text{PLocation} \}$ $\xrightarrow[\text{Rule}]{\text{Augmentation}}$

$\{ \text{Pnumber} \} \rightarrow \{ \text{Pnumber}, \text{Pname}, \text{PLocation} \}$

• $\{ \text{ssN} \} \rightarrow \{ \text{ssN}, \text{Ename} \}$ $\xrightarrow[\text{dependencies}]{\text{functional}}$

$\{ \text{ssN}, \text{Pnumber} \} \rightarrow \{ \text{ssN}, \text{Ename} \} \dots A$

• $\{ \text{Pnumber} \} \rightarrow \{ \text{Pnumber}, \text{Pname}, \text{PLocation} \}$ $\xrightarrow[\text{dependencies}]{\text{functional}}$

$\{ \text{ssN}, \text{Pnumber} \} \rightarrow \{ \text{Pnumber}, \text{Pname}, \text{PLocation} \} \dots B$

• $A, B \xrightarrow[\text{Rule}]{\text{Union}}$

$\{ \text{ssN}, \text{Pnumber} \} \rightarrow \{ \text{ssN}, \text{Pnumber}, \text{Ename}, \text{Pname}, \text{PLocation}, \text{Hours} \}$

Normalization

- ◆ Normalization
 - Proposed by Codd
 - Proceeds in a **top-down fashion** by evaluating each relation against the criteria for normal forms & decomposing relations as necessary
 - Relation design by analysis
- ◆ Normal forms
 - 1NF (First normal form)
 - 2NF (Second normal form)
 - 3NF (Third normal form)
 - BCNF (Boyce-Codd normal form)
 - 4NF (Fourth normal form): multi-value dependency
 - 5NF (Fifth normal form): join-dependency

functional dependencies
among attributes