

## Functional Dependencies - II

# Example Schema

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
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

- What is this table about?
- Your intuition: is it well designed?

# Normalization

- Theory and process by which to evaluate and improve relational database design
- Typically divide larger tables into smaller, less redundant tables
  - Focus now on correctness (we'll return to the possibility of “denormalization” in physical design for improving efficiency)

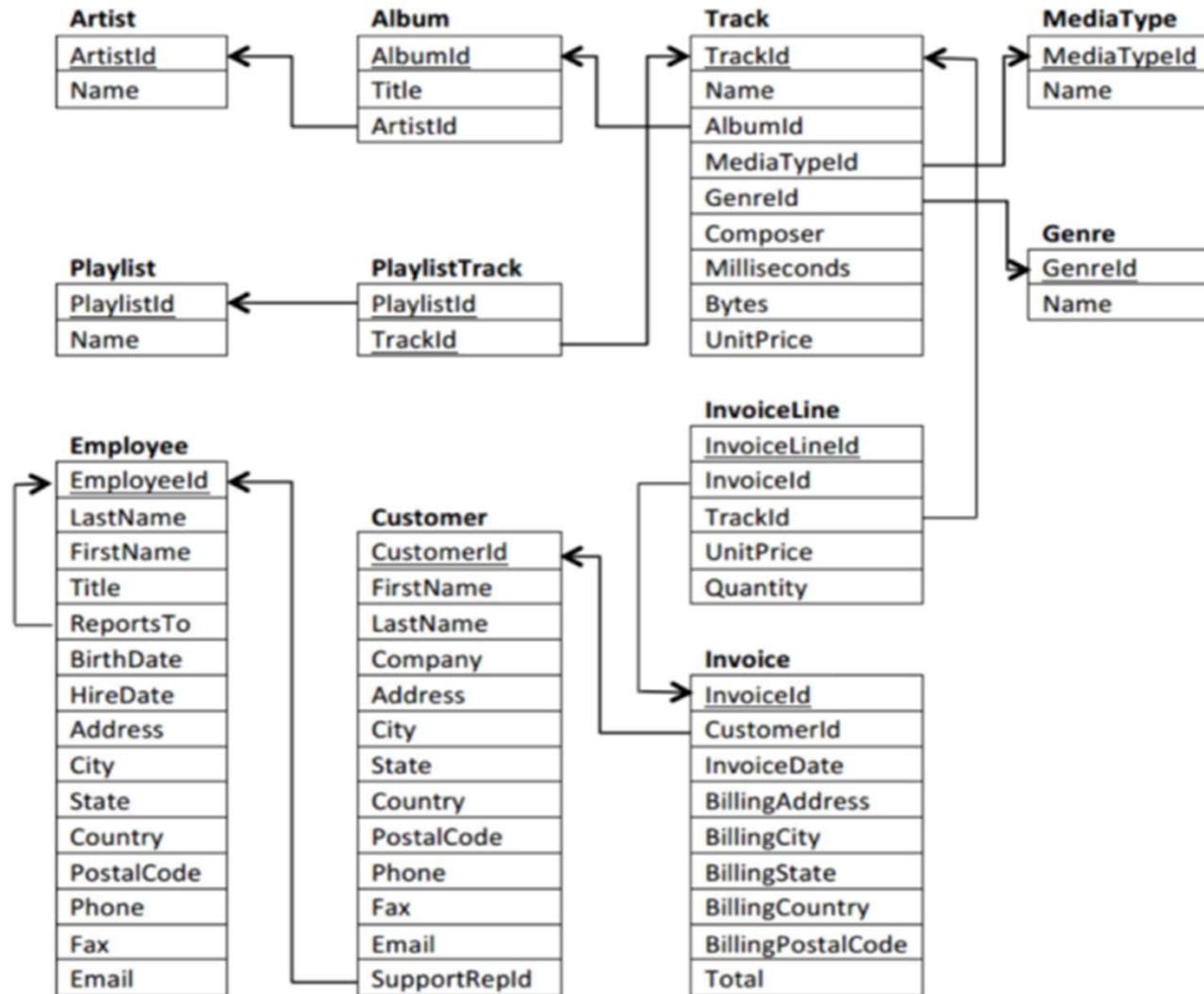
# Objectives of Normalization

- Make the schema informative
- Minimize information duplication
- Avoid **modification anomalies**
- Disallow **spurious tuples**

# Make the Schema Informative

- Design a relational schema so that it is easy to explain its meaning
- Do **not** combine attributes from multiple entity types and relationship types into a single relation
  - Can make explaining/understanding difficult
- Ideally, tables & constraints mirror real-world concepts & relationships

# Tracks from the Past



# Example Schema

What is this table about?

- Employees? Departments?

EMP\_DEPT

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
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

# Minimize Information Duplication

Redundancy

EMP\_DEPT

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
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



# Avoid **Modification Anomalies**

An undesired side-effect resulting from an attempt to modify a table (that has not been sufficiently normalized)

# Insertion Anomaly

Difficult or impossible to insert a new row

- Create the new “Marketing” department

EMP\_DEPT

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
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

# Update Anomaly

Updates may result in logical inconsistencies

- Change Ramesh's department name to R&D

EMP\_DEPT

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
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

# Deletion Anomaly

Deletion of data representing certain facts necessitates deletion of data representing completely different facts

- Delete James E. Borg

EMP\_DEPT

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
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

# Disallow Spurious Tuples

*Avoid relational design that matches attributes across relations that are not (foreign key, primary key) combinations because joining on such attributes may produce invalid tuples*

***Spurious tuples are extra, incorrect rows that appear when you join decomposed tables in a lossy decomposition. They represent data combinations that **did not exist in the original relation**, leading to inaccurate results.***

# Game Plan

Build up to a set of “tests” (*Normal Forms*) that indicate cumulatively improving degrees of design quality

Redundancy

EMP_DEPT						
Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
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 Fire Oak, 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

EMPLOYEE

Ename	<u>Ssn</u>	Bdate	Address	Dnumber
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

DEPT\_LOCATIONS

<u>Dnumber</u>	<u>Dlocation</u>
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

# Detour: Formalization

- We need a way of understanding how data in our tables depend on each other (termed: functional dependencies)
- With just this information, we can identify (candidate) keys and analyze/improve schema design
  - Even without knowing what the data is about!

# FD Example (1)

	StudentID	Year	Class	Instructor
$t_1$	1	Sophomore	CS3200	Rachlin
$t_2$	2	Sophomore	DS2500	Rachlin
$t_3$	3	Junior	CS3200	Rachlin
$t_4$	3	Junior	DS2500	Rachlin
$t_5$	2	Sophomore	CS3200	Derbinsky
$t_6$	4	Sophomore	CS3200	Derbinsky

What FDs hold in the current state of this relation?

$\{StudentID\} \rightarrow \{Year\}$

$\{StudentID, Class\} \rightarrow \{Instructor\}$



# FDs & Keys

- One cannot determine whether FDs hold for *all* relation states unless the meaning of and relationships among the attributes are known

# FD Example (2)

	StudentID	Year	Class	Instructor
$t_1$	1	Sophomore	CS3200	Rachlin
$t_2$	2	Sophomore	DS2500	Rachlin
$t_3$	3	Junior	CS3200	Rachlin
$t_4$	3	Junior	DS2500	Rachlin
$t_5$	2	Sophomore	CS3200	Derbinsky
$t_6$	4	Sophomore	CS3200	Derbinsky

What FDs hold in the current state of this relation?

$\{StudentID\} \rightarrow \{Year\}$

$\{StudentID, Class\} \rightarrow \{Instructor\}$

Multiple sections?  
Co-instruction?  
Re-takes?

# FDs & Keys

- One cannot determine whether FDs hold for *all* relation states unless the meaning of and relationships among the attributes are known
  - These are the “data dependencies” foreshadowed in Lecture 2 (Relational Model)
  - If you *do* have this domain knowledge, it is possible to identify candidate keys (minimal subsets of attributes that FD all attributes)

# FD Example (3)

	StudentID	Year	Class	Instructor
$t_1$	1	Sophomore	CS3200	Rachlin
$t_2$	2	Sophomore	DS2500	Rachlin
$t_3$	3	Junior	CS3200	Rachlin
$t_4$	3	Junior	DS2500	Rachlin
$t_5$	2	Sophomore	CS3200	Derbinsky
$t_6$	4	Sophomore	CS3200	Derbinsky

$\{StudentID\} \rightarrow \{Year\}$

$\{StudentID, Class\} \rightarrow \{Instructor\}$

**Key(s):**  $\{StudentID, Class\}$

- Every student is classified as either a Freshman, Sophomore, Junior, or Senior.
- Students can take only a single section of a class, taught by a single instructor.

# FDs & Keys

- One cannot determine whether FDs hold for *all* relation states unless the meaning of and relationships among the attributes are known
  - If you *do* have this domain knowledge, it is possible to identify candidate keys (minimal subsets of attributes that FD all attributes)
- One can state an FD does *not* hold given a relation state by identifying violating tuple(s)

# FD Example (4)

	StudentID	Year	Class	Instructor
$t_1$	1	Sophomore	CS3200	Rachlin
$t_2$	2	Sophomore	DS2500	Rachlin
$t_3$	3	Junior	CS3200	Rachlin
$t_4$	3	Junior	DS2500	Rachlin
$t_5$	2	Sophomore	CS3200	Derbinsky
$t_6$	4	Sophomore	CS3200	Derbinsky

$\{StudentID\} \not\rightarrow \{Instructor\}$

$\{StudentID\} \not\rightarrow \{Class\}$

$\{Year\} \not\rightarrow \{StudentID\}$

$\{Year\} \not\rightarrow \{Instructor\}$

$\{Year\} \not\rightarrow \{Class\}$

$\{Class\} \not\rightarrow \{Year\}$

$\{Class\} \not\rightarrow \{StudentID\}$

$\{Class\} \not\rightarrow \{Instructor\}$

$\{Instructor\} \not\rightarrow \{Class\}$

$\{Instructor\} \not\rightarrow \{Year\}$

$\{Instructor\} \not\rightarrow \{StudentID\}$

# FD Example (5)

	StudentID	Year	Class	Instructor
$t_1$	1	Sophomore	CS3200	Rachlin
$t_2$	2	Sophomore	DS2500	Rachlin
$t_3$	3	Junior	CS3200	Rachlin
$t_4$	3	Junior	DS2500	Rachlin
$t_5$	2	Sophomore	CS3200	Derbinsky
$t_6$	4	Sophomore	CS3200	Derbinsky

$\{StudentID, Instructor\} \not\rightarrow \{Class\}$

$\{Year, Class\} \not\rightarrow \{Instructor\}$

$\{Year, Class\} \not\rightarrow \{StudentID\}$

$\{Class, Instructor\} \not\rightarrow \{StudentID\}$

$\{Class, Instructor\} \not\rightarrow \{Year\}$

$\{Year, Class, Instructor\} \not\rightarrow \{StudentID\}$