The Relational Model of Data

5DV119 — Introduction to Database Management

Umeå University
Department of Computing Science

Stephen J. Hegner
hegner@cs.umu.se
http://www.cs.umu.se/~hegner

The Main Idea

- In the relational model, the data are stored in *relations*, or *tables*.
- An example is shown on the next slide.
  - The *relation names* are shown with a blue background.
  - The *attribute names* are shown with a lime background.
  - The *tuples* are shown with a beige background.
  - The *primary keys* are underlined.

Databases are sets: For each relation, the set of rows consists of distinct elements.

- Duplicate rows are not allowed.
- The order in which the rows are displayed is of no formal consequence.

The Company Schema with a Simple Database Instance

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	Zeyala	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_Location

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
453453453	2	20
333445555	2	10
333445555	3	10
333445555	10	10
333445555	20	10
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

Project

PName	PNumber	PLocation	DNumber
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Administration	30	Stafford	4

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	1942-02-08	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

Attributes and Domains

- The label of a column is called an *attribute*.
- With each attribute A is associated a set Dom(A) of *domain values* for A.

Example: Dom(SSN) might be the set of all strings of digits of length exactly nine.

- The *field* identified by attribute A of a tuple must have values taken from Dom(A).
- The only exception is that in some cases the special value NULL may be used.
  - More later on null values.

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	Zeyala	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

Relation Schemes

- A *relation scheme* (or *relation schema*) is a type definition for a table.
  - Formally, it is a (usually nonempty) finite list of attributes.
  - In the classical literature, it is often taken to be a *set* of attributes.
  - The order of the attributes is of no conceptual importance.
- The six relation schemes of the company database are shown below.
- The relation name is the same as the scheme name.

Employee

FName	MInit	LName	SSN	BDate	Address	Sex	Salary	Super_SSN	DNo

Department

DName	DNumber	Mgr_SSN	Mgr_Start_Date

Dept_Location

DNumber	DLocation

Project

PName	PNumber	PLocation	DNumber

Works_On

ESSN	PNo	Hours

Dependent

ESSN	Dependent_Name	Sex	BDate	Relationship

Tuples

Context: $R = (A_1, A_2, \ldots, A_k)$ a relation scheme.

Tuple: A row of data for R is called a *tuple* for R:

$$t = (t_1, t_2, \ldots, t_k) \text{ with } t_i \in \mathrm{Dom}(A_i) \cup \{\mathrm{NULL}\} \text{ for each } i.$$

Example: $t_{\mathrm{Joyce}} =$ ('Joyce','A','English','453453453','1972-07-31', '5631 Rice, Houston, TX','F',25000,'333445555',4)

Projection of a tuple: For $\mathbf{A} = (A_{i_1}, A_{i_2}, \ldots, A_{i_\ell})$ a nonempty subsequence of R (*i.e.*, $1 \leq i_1 < i_2 < \ldots < i_\ell \leq k$) and $t = (t_1, t_2, \ldots, t_k)$ a tuple over R, define the *projection* of t onto $\mathbf{A}$ to be $t[\mathbf{A}] = (t_{i_1}, t_{i_2}, \ldots, t_{i_\ell})$

Example:

t_{Joyce}[LName,FName,SSN,Sex] = ('Joyce','English','453453453','F')

Notation: When abstract schemes are considered for example, $\mathbf{A} = (A_{i_1}, A_{i_2}, \ldots, A_{i_\ell})$, commas may be dropped in attribute lists.

Example: $t[A_1, A_3, A_7] = t[A_1A_3A_7]$.

Notational convention: For a *subset* $S = \{A_{i_1}, A_{i_2}, \ldots, A_{i_\ell}\} \subseteq \{A_1, A_2, \ldots, A_k\}$, write $t[S]$ for $t[A_{i_1}, A_{i_2}, \ldots, A_{i_\ell}]$.

Integrity Constraints

- The allowable databases must satisfy certain *integrity constraints*.

Examples:

- No two employees may have the same SSN.
- The MgrSSN of a department must be the SSN of some employee.
- The SuperSSN of an employee must either be the SSN of some employee or else NULL.
- The LName of an employee may not be NULL.

Employee

FName	MInit	LName	SSN	BDate	Address	Sex	Salary	Super_SSN	DNo

Department

DName	DNumber	Mgr_SSN	Mgr_Start_Date

Dept_Location

DNumber	DLocation

Project

PName	PNumber	PLocation	DNumber

Works_On

ESSN	PNo	Hours

Dependent

ESSN	Dependent_Name	Sex	BDate	Relationship

Flavors of Keys for Instances for Relation Schemes

Context: $R = (A_1, A_2, \ldots, A_k)$ a relation scheme.

Notation: Write Attrset(R) for $\{A_1, A_2, \ldots, A_k\}$.

- An *instance* M_R for R is a set of tuples for R.
- $S \subseteq$ Attrset(R) is a *superkey* for the instance M_R if for any two $t, t' \in M_R$, if $t[S] = t'[S]$ then $t = t'$.
- $K \subseteq$ Attrset(R) is a *key* for the instance M_R if it is a superkey for M_R, and no proper subset $K' \subsetneq K$ is a superkey.

Examples: Each of the attributes (as a singleton set) FName, MInit, LName, SSN, BDate, Address is a key for the instance below, as is {Sex, Salary, Super_SSN}.

- Any superset of these is a superkey.

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	Zeyala	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

Keys Constraints on Relation Schemes

Context: $R = (A_1, A_2, \ldots, A_k)$ a relation scheme.

Key constraint The *key constraint* for R defined by $K \subseteq \text{Attrset}(R)$ is the requirement that all allowed instances M_R for R have K as a key.

Primary key: Every relation scheme which is part of a relational schema must have a distinguished key, called the *primary key*, which defines a key constraint.

- Whether or not a given K defines a key constraint is a modelling decision, determined by the context being modelled.
- In drawing a relation scheme, the primary key is usually underlined.
- In the Employee relation, with the "obvious" semantics, SSN is the only possible key, and so is the primary key.

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	Zeyala	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

Multiple Candidate Keys

- It is possible for a relation scheme to have several natural key constraints.

Example: The Department relation of the Company database has both DNumber and DName as possible keys.

Example: The Project relation of the Company database has both PNumber and PName as possible keys.

- Each of these is called a *candidate key*.
- DNumber and PNumber are chosen as the primary keys, but these are somewhat arbitrary design decisions.

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

Project

PName	PNumber	PLocation	DNumber
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Administration	30	Stafford	4

Another Example of Multiple Candidate Keys

- In the database in which the grades for this course are maintained, there is a relation which contains identification information about students.

Student

Ident	PersonNr	Name

Attributes:

Ident: The user-id of the student on the departmental computing systems.

PersonNr: The Swedish identification number of the student.

Name: The name of the student.

- Both Ident and PersonNr are candidate keys.
- Ident is chosen because it is more stable.
- International students are sometimes given a temporary PersonNr by the university, which changes once the permanent one is known.
- The Ident does not usually change.

Keys with Several Attributes

- It is not always the case that a key consists of a single attribute.
- Three of the relations in the Company schema have primary keys consisting of two attributes.

Employee

FName	MInit	LName	SSN	BDate	Address	Sex	Salary	Super_SSN	DNo

Department

DName	DNumber	Mgr_SSN	Mgr_Start_Date

Dept_Location

DNumber	DLocation

Project

PName	PNumber	PLocation	DNumber

Works_On

ESSN	PNo	Hours

Dependent

ESSN	Dependent_Name	Sex	BDate	Relationship

Foreign-Key Dependencies

Example: The manager of a department must be an employee.

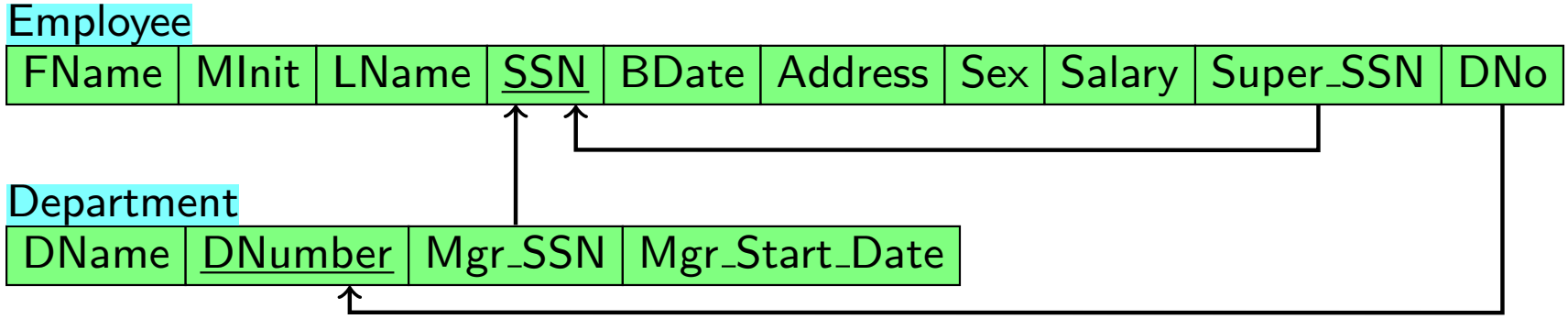
- Formally, every value for a Mgr_SSN in the Department relation must be a value for the primary key SSN of Employee.

Foreign key: Mgr_SSN is said to be a *foreign key* of the Department relation which *references* the Employee Relation.

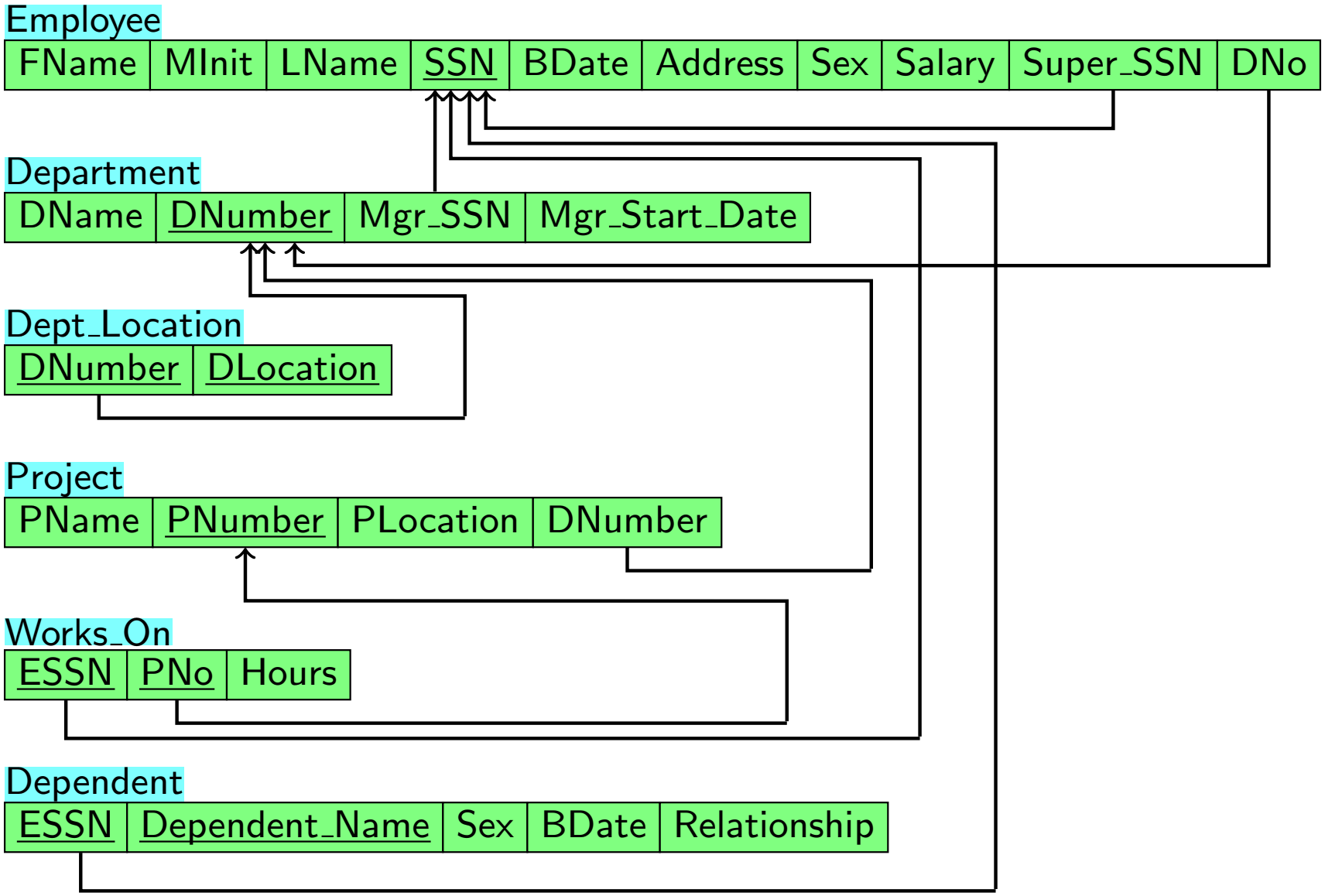
- The associated constraints are called *foreign-key constraints* or *referential integrity constraints*.

Notation: A foreign key relationship is represented by drawing an arrow from the foreign key to the associated primary key.

- There are other foreign-key dependencies on these two relation schemes.
  - Note that both the foreign key and the referenced primary key may be in the same relation.



All Foreign-Key Constraints on the Company Schema



Foreign-Key Constraints and Null Values

- In general, a foreign key may be null.

Example: Not every employee can have a supervisor.

  - Someone must be the big boss.

- However, some foreign keys must not be null.

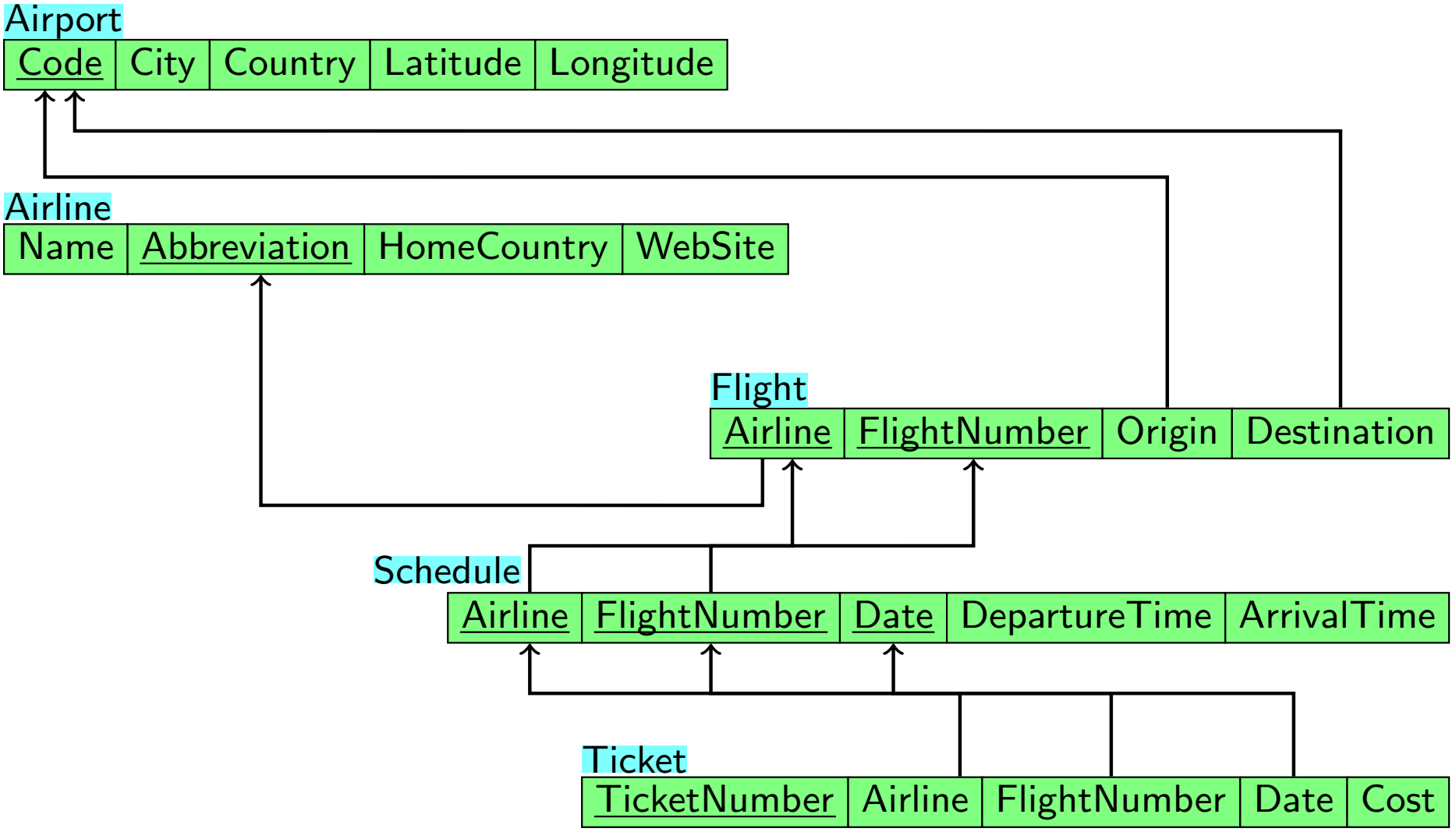
Example: Every dependent must be associated with an employee.

- Whether or not an attribute may be null is determined in the formal schema definition, to be studied shortly.

Composite Foreign Keys

- It is important to see that a foreign key may consist of more than one attribute.
- This happens precisely when the associated primary key consists of more than one attribute.
  - A foreign key is *always* associated with the primary key of another (possibly the same) relation.
- An air-travel schema is shown on the next slide.
- Note in particular that there are two foreign keys with multiple attributes.
  - (Airline,FlightNumber) is a foreign key of the Schedule relation which references the Flight relation.
  - (Airline,FlightNumber,Date) is a foreign key of the Ticket relation which references the Schedule relation.
- Note also that the graphical notation collapses the multiple lines into one, to indicate that it is a composite foreign key.
- A composite foreign key is not the same as simply requiring a subset relationship on each attribute individually.

The Air-Travel Schema



Relational Database Schemata and Relational Databases

Relational Database Schema: $\mathbf{R}$ defined by:

- A set Schemes($\mathbf{R}$) $= \{R_1, R_2, \ldots, R_k\}$ of relation schemes.
- A set Constr($\mathbf{R}$) of constraints (in particular key and foreign key) on the members of Schemes($\mathbf{R}$).

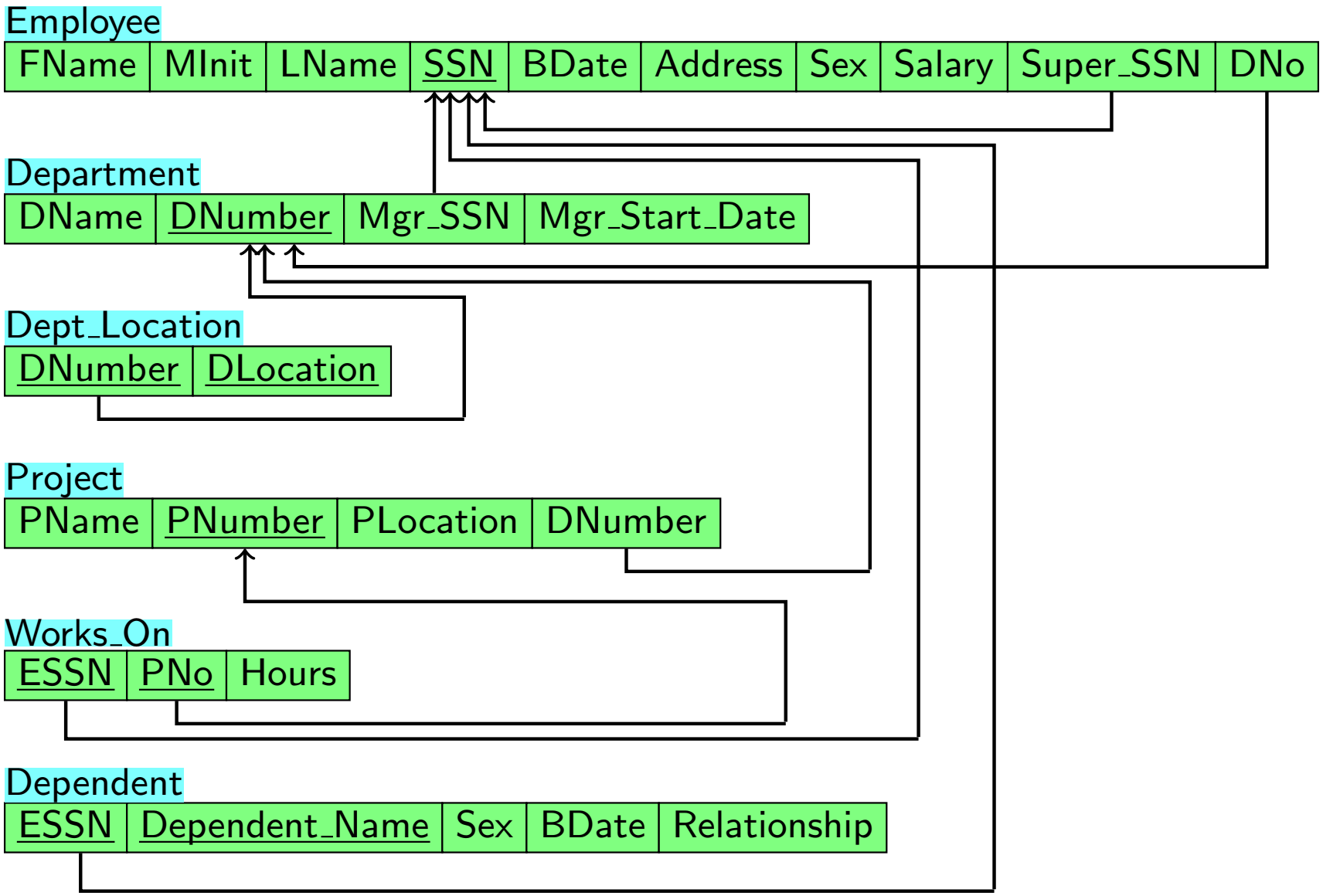
Relational Database: on $\mathbf{R}$ is given by a set

$$M = \{M_R \mid R \in \text{Schemes}(\mathbf{R}) \text{ and } M_R \text{ is an instance for R}\}.$$

Legal (Relational) Database: on $\mathbf{R}$ consists of those databases which satisfy the constraints in Constr($\mathbf{R}$).

- The schema is generally static, and is changed only rarely.
  - Changing the schema may require extensive data reorganization.
- Changing (*updating*) the data in the database, on the other hand, is a central operation in database management.

Graphical Representation of a Relational Database Schema



- Note that only primary- and foreign-key constraints are shown.

A Relational Database for the Schema

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	Zeyala	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_Location

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
453453453	2	20
333445555	2	10
333445555	3	10
333445555	10	10
333445555	20	10
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

Project

PName	PNumber	PLocation	DNumber
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Administration	30	Stafford	4

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	1942-02-08	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

Null Values

- A special value NULL is allowed instead of a normal domain value in certain circumstances.
- The semantics of NULL is unfortunately not specified.
  - The semantics must be defined by usage convention.

Example: Suppose that the attribute *Telephone* is added to the employee relation.

- There are at least three possible interpretations of a null value:

  Value not known: The employee has a telephone but the number is not in the database.

  No value: The employee has no telephone (but could have one).

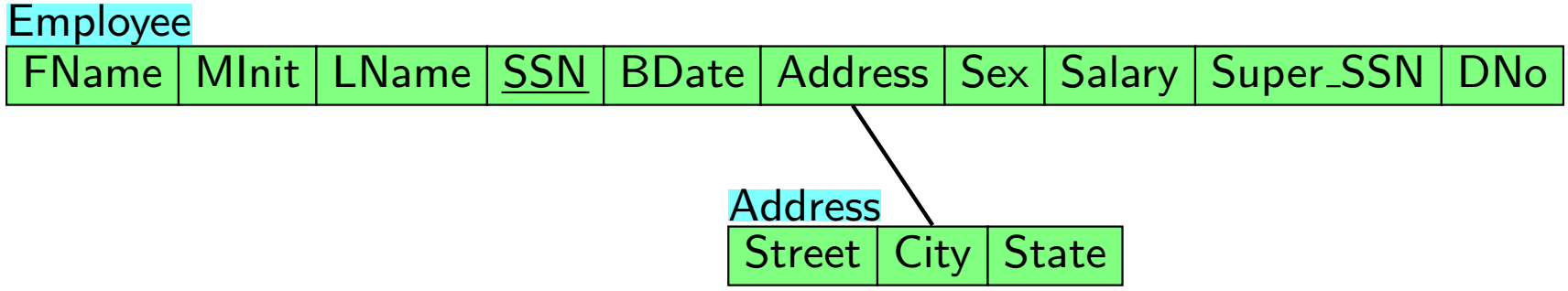
  Inappropriate attribute: The position of the employee (*e.g.*, custodial) does not involve having a telephone.

- Thus, the meaning of null values can be ambiguous.

Question: What do the null values in the example Company database mean?

First Normal Form

- It would be desirable to be able to decompose attributes such as Address into tuples of subattributes, as illustrated below.
- With the classical relational model, this is not possible.
- In so-called *first normal form (1NF)*, all domains are atomic.
- In the Employee relation, the attribute Address could be replaced by the three attributes Street, City, and State, but the ability to refer to Address as their composite would be lost.
- A representation as illustrated below is available in some relational systems as an *object-relational* extension.
- It has even become part of the latest SQL standard.
- However, support is still far from universal or uniform.
- Object-relational extensions will not be considered in this course.



Update Operations

- There are three main classes of update operations:
  - Insert: Add some new tuples to the database.
  - Delete: Remove some tuples from the database.
  - Modify: Change some fields of existing tuples.
- A principal aspect of support for update operations is to ensure that the integrity constraints remain satisfied.
- There are two main strategies:
  - Restrict: If a proposed update violates the integrity constraints, reject it.
  - Cascade: If a proposed update violates the integrity constraints, make additional updates automatically to satisfy them.

Example of cascading: If an employee is deleted, remove also all associated entries in the Works_On and Dependent relations.

- It is probably not a good idea to cascade to the Department or Employee relations!!

Remark on terminology: *Modify* is called *Update* in SQL.

- This leads to terminology overload which can be confusing and must be resolved carefully.

Transactions

- Most DBMSs allow simultaneous operations by several users/processes.
- To avoid conflict, the notion of a *transaction* has evolved.
- The idea is that distinct transactions should not interfere with each other.
- An entire theory and practice of database transactions has evolved.
- Unfortunately, it is not possible to cover the topic in any detail in this course.
    - Transactions are covered in detail in the course 5DV120.
- However, it is still useful to be aware that transaction support is central to any modern DBMS.

Authorization

- It is neither wise nor practical to give all users complete access to the entire database.
- Rather, users are given only specific access rights which are necessary to carry out their tasks.

  Example: In the database of your bank, you have access only to information concerning your accounts.

Authorization mechanism: An *authorization mechanism* provides support for two fundamental actions:

Grant: This operation allows a given user to give access privileges to other users on specific database objects, including the privilege to grant to others.

Revoke: This operation allows a given user to take back privileges which that user had previously granted.

- Modern database systems typically support a rather extensive authorization mechanism which is part of SQL.
  - This mechanism will be studied near the end of the course.