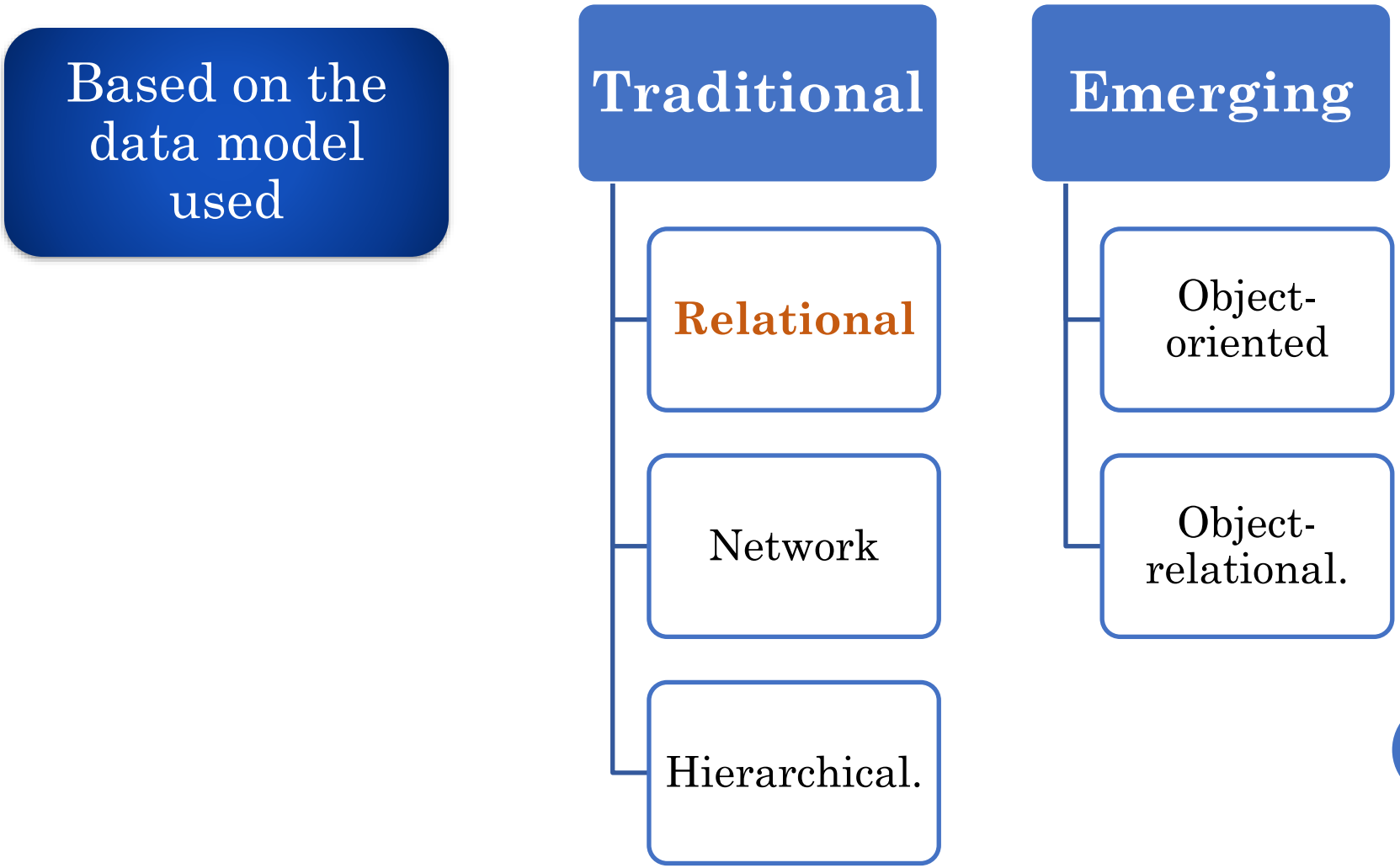
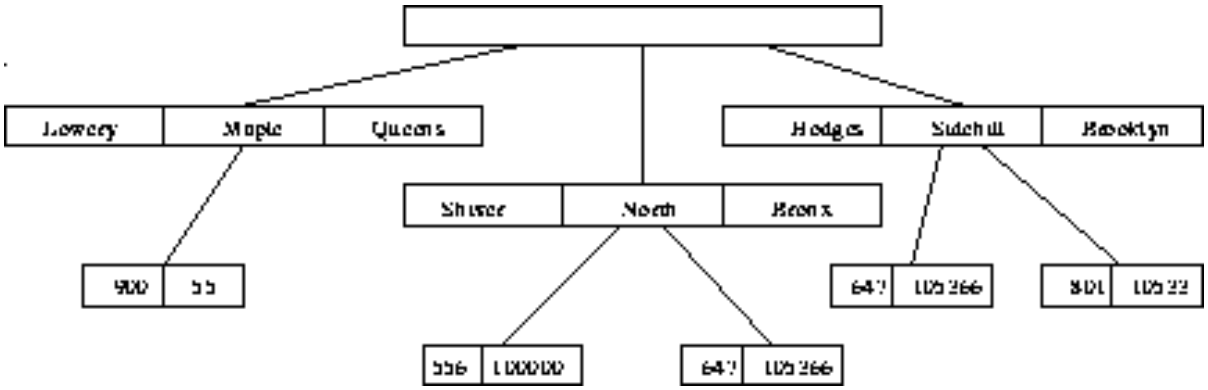
Classification of DBMSs



Hierarchical Model

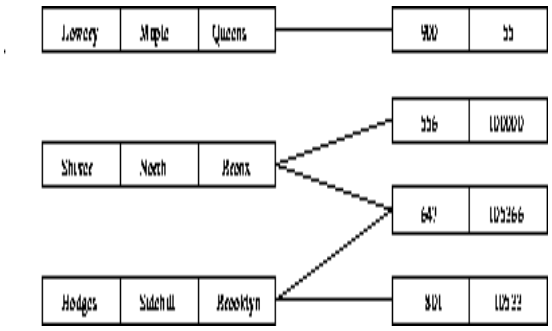
- Organization of the records is as a collection of **trees**
- A child can have only one parent, but a parent can have many children.

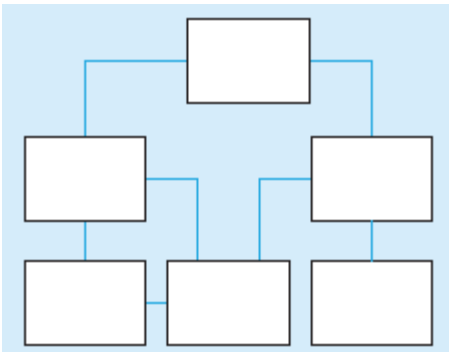


- Drawback- Data redundancy
  - Cannot handle many-to-many relationship
  - Example: employee with many supervisors

Network Model

- Data are represented by collections of records.
- Relationships among data are represented by links.
- Organization is that of an **arbitrary graph**.
- Eliminates hierarchical model problem issue
  - Allow many to many relationship





- Limitations: difficult to query data from multiple tables and issues in indexing scheme

Relational Model

Relational databases use **tables** to store information.

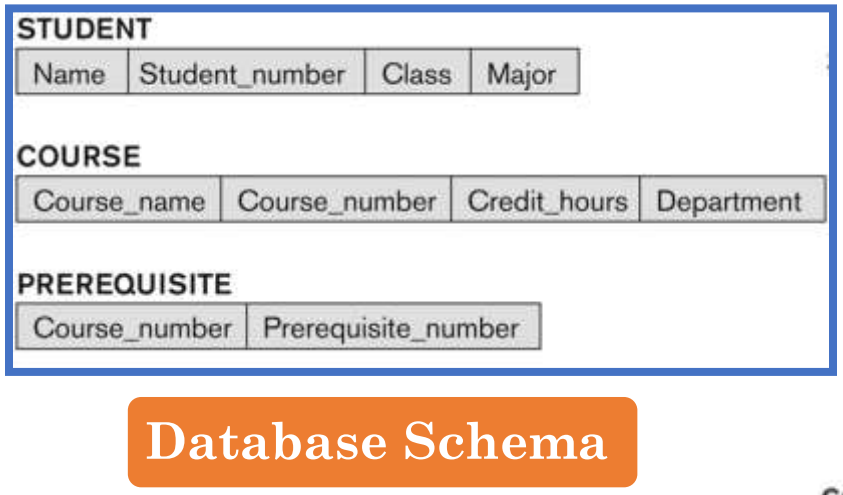
EMPLOYEE

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

Primary Key: uniquely identifies each record in the table

DB Schemas And State

- **Schema:** Description of a database *(in a given data model)*
- **State:** Refers to the content of a database at a moment in time.



Database Schema

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

Database State

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

DB Schemas And State

- **Schema** changes very infrequently, whereas **state** changes every time the database is updated.
- **Schema** is also called **intension**, whereas **state** is called **extension**.

STUDENT

Name	Student_number	Class	Major

COURSE

Course_name	Course_number	Credit_hours	Department

PREREQUISITE

Course_number	Prerequisite_number

Database Schema

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

Database State

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

Schema

- **Schema** of a Relation: R (A_1, A_2, ...A_n)

STUDENT(*Name, Ssn, Phone, Address, Age, Gpa)*

- Each attribute has a **domain** or a set of valid values.
- **Example:**
  - Ssn: set of valid 9 digit social security numbers.
  - Phone: the set of 10 digit phone numbers valid in the United States.

Domain is Atomic (indivisible) or NULL

- Integers and strings,
- Record structures: set, list, array … ??

State

- Relation State **s(R)** is set of tuples
- Each **tuple** t is an ordered list of n values

$$t = <v_1, v_2, \ldots, v_n>$$

where each value v_i, $1 \leq i \leq n$, is an element of $dom(A_i)$ or is NULL

Ordering of column in R is important

However, No ordering of tuples(rows) in r(R)

WHY ??

STUDENT

Name	Ssn	Home_phone	Address	Office_phone	Age	Gpa
Dick Davidson	422-11-2320	NULL	3452 Elgin Road	(817)749-1253	25	3.53
Barbara Benson	533-69-1238	(817)839-8461	7384 Fontana Lane	NULL	19	3.25
Rohan Panchal	489-22-1100	(817)376-9821	265 Lark Lane	(817)749-6492	28	3.93
Chung-cha Kim	381-62-1245	(817)375-4409	125 Kirby Road	NULL	18	2.89
Benjamin Bayer	305-61-2435	(817)373-1616	2918 Bluebonnet Lane	NULL	19	3.21

Formal Definitions: Relation

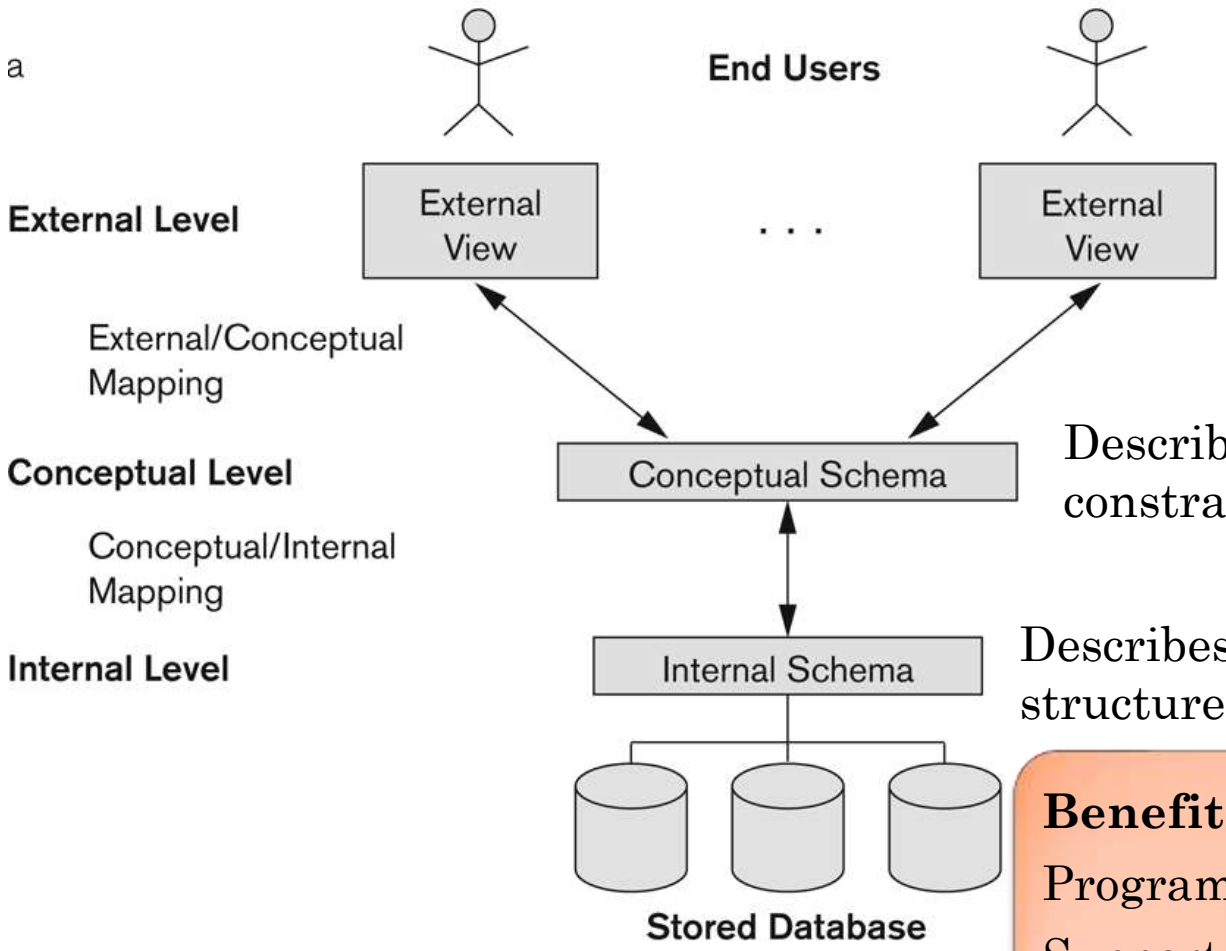
- A relation $r(R)$ is a **mathematical relation** of degree n on the domains $\text{dom}(A_1),\ldots, \text{dom}(A_n)$.
  - **s(R)** is a **subset of Cartesian product** of the domains that define R

$$s(R) \subseteq (\text{dom}(A_1) \times \text{dom}(A_2) \times \ldots \times \text{dom}(A_n))$$

- Example:
  - Let S1 = {0,1} and S2 = {a,b,c}
  - s(R) $\subset$ S1 X S2
  - s(R) = {<0,a> , <0,b> , <1,c> } is one possible "state"

The total no of possible tuples that can ever exist in any relation state $s(R)$ is $|\text{dom}(A_1)| \times |\text{dom}(A_2)| \times \ldots \times |\text{dom}(A_n)|$

Three-Schema Architecture



Describes the various user views

Describes the structure and constraints for the database

Describes physical storage structures and access paths.

Benefits:

Program data independence

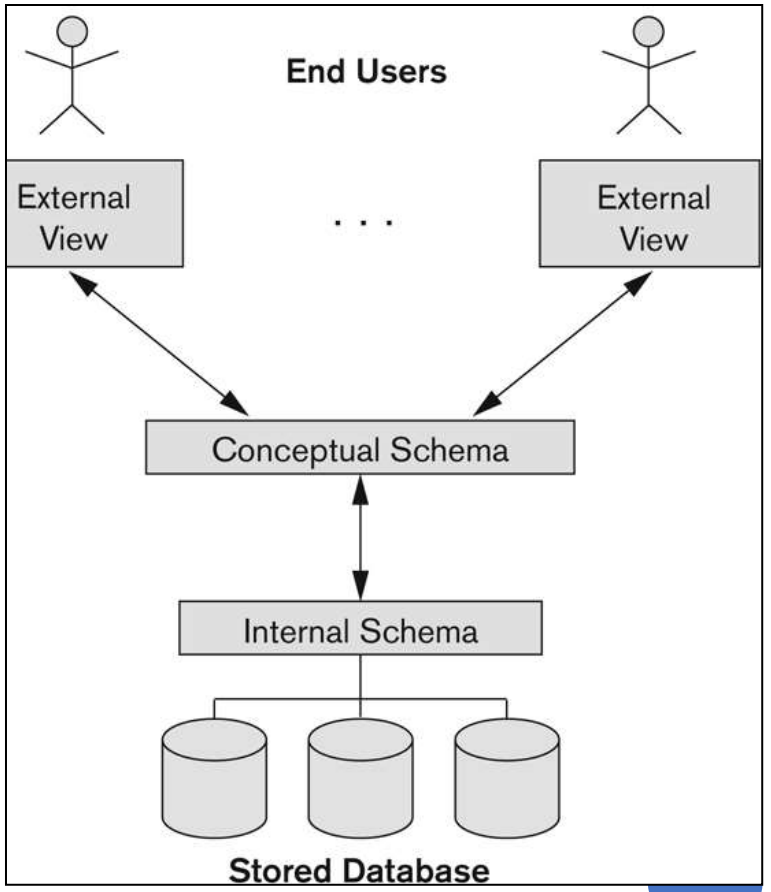
Support multiple views of the data.

Data Independence

Concept: Applications do not need to worry about *how the data is structured and stored*

Physical Data Independence

- ***=> change the internal schema without changing the conceptual schema.***
- protection from *physical layout changes*
- For example: the internal schema may be changed
  - when certain file structures are reorganized or
  - new indexes are created to improve database performance

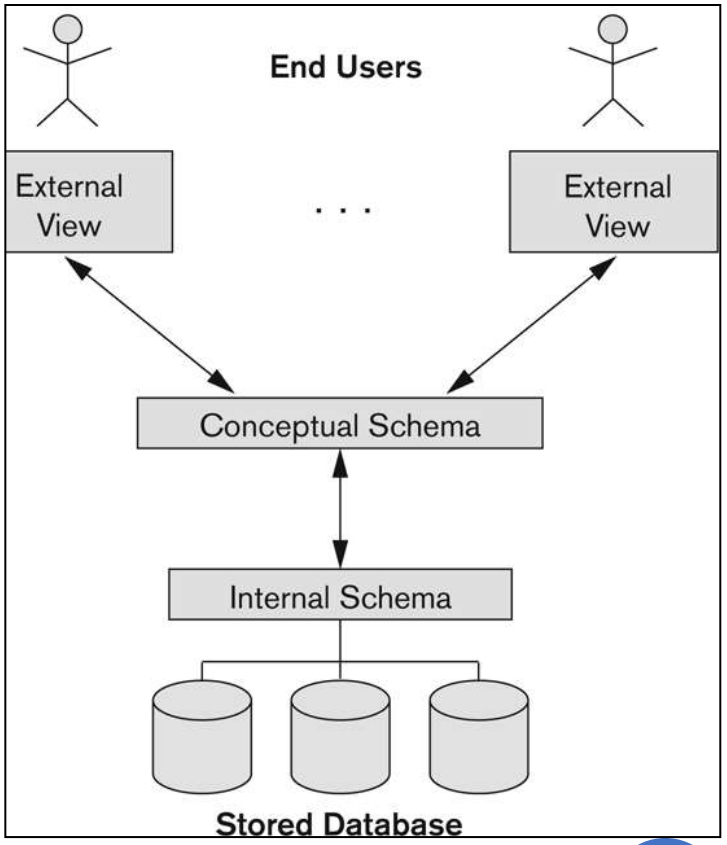


One of the most important benefits of using a DBMS!

Data Independence

Logical Data Independence

- ***=> change the conceptual schema without changing the external schemas and application programs.***
- Protect from changes in the *logical structure of the data*
- *Only the mappings need to be changed in DBMS*



It is harder to achieve because it allows structural and constraint changes without affecting application programs

Logical Data Independence

GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

GRADE_REPORT

Student_number	Student_name	Section_identifier	Course_number	Grade
17	Smith	112	MATH2410	B
17	Smith	119	CS1310	C
8	Brown	85	MATH2410	A
8	Brown	92	CS1310	A
8	Brown	102	CS3320	B
8	Brown	135	CS3380	A

Informal Terms		Formal Terms
Table		Relation
Column		Attribute/Domain
Row		Tuple
Values in a column		Domain
Table Definition		Schema of a Relation
Populated Table		State (Extension)

Key Constraint

- Relation is a **set** of tuples. Thus, all tuples must be distinct as all elements in the set are distinct

- **Example 1:** Student Relation

Student(Name, SSN, Address, Phone, Major, Age, Gpa)

  - **KEY is** SSN, as it uniquely identify the student.

- **Example 2: Department** Relation

Department(Dname, Dnumber, MgrSSN, Mgrstrdate)

  - KEY is Dnumber
  - Candidate Key Dname

Example of a University database

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

SESSION

Session-id	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

ENROLLMENT

Student_number	Session-id	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

Key Constraint

- **Key**:
  - The **"minimal"** set of attributes that uniquely determine the tuple (row)
  - Key is a set from which we cannot remove any attribute and still have the uniqueness constraint hold.

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

minimum and minimal -- session id
Minimal --- Cno,sem,year

SESSION

Session-id	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

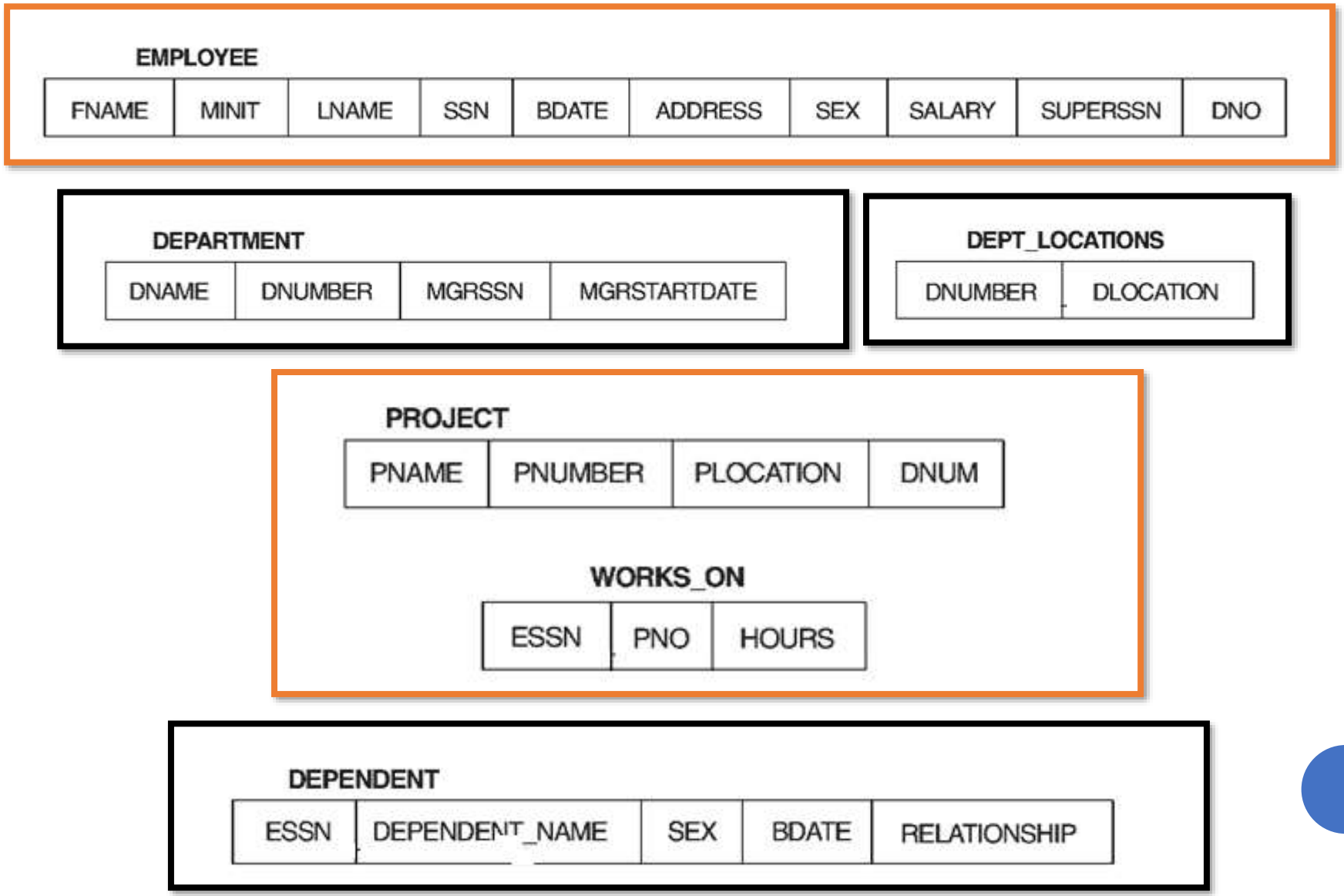
Key Constraint

- **CAR relation schema**

CAR	LicenseNumber	EngineSerialNumber	Make	Model	Year
	Texas ABC-739	A69352	Ford	Mustang	96
	Florida TVP-347	B43696	Oldsmobile	Cutlass	99
	New York MPO-22	X83554	Oldsmobile	Delta	95
	California 432-TFY	C43742	Mercedes	190-D	93
	California RSK-629	Y82935	Toyota	Camry	98
	Texas RSK-629	U028365	Jaguar	XJS	98

- **Key**
  - Key1 = {LicenseNo}
  - Key2 = {Engine_SerialNo},

Key Constraint



Key Constraint

EMPLOYEE

FNAME	MINIT	LNAME	SSN	BDATE	ADDRESS	SEX	SALARY	SUPERSSN	DNO

DEPARTMENT

DNAME	DNUMBER	MGRSSN	MGRSTARTDATE

DEPT_LOCATIONS

DNUMBER	DLOCATION

PROJECT

PNAME	PNUMBER	PLOCATION	DNUM

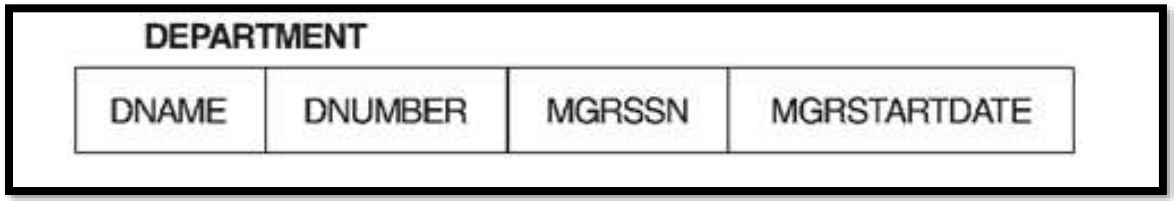
WORKS_ON

ESSN	PNO	HOURS

DEPENDENT

ESSN	DEPENDENT_NAME	SEX	BDATE	RELATIONSHIP

Entity Integrity



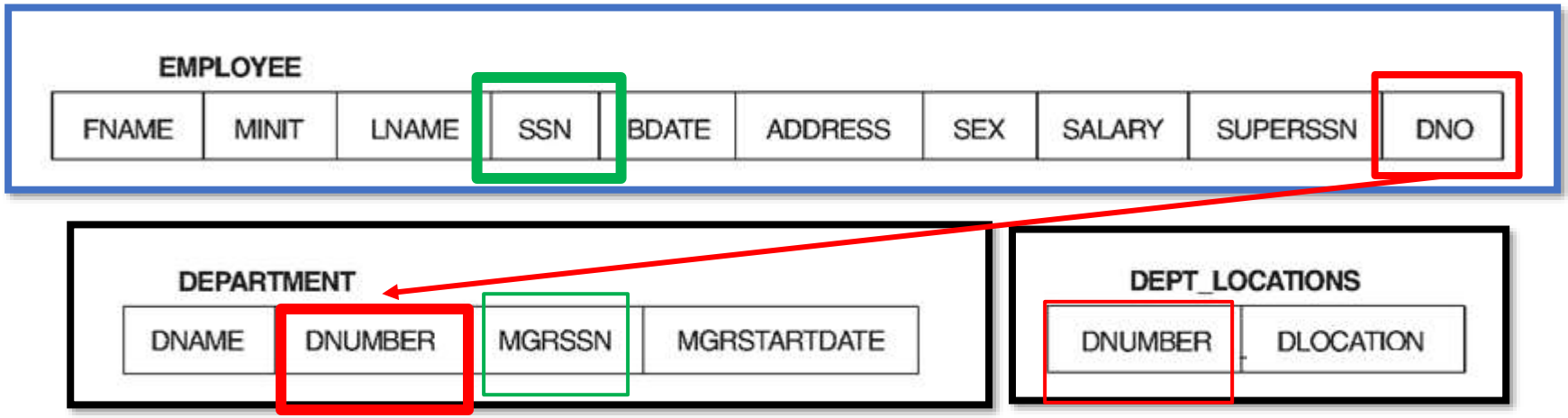
Can DNUMBER be NULL?

Primary key is used to *identify* the individual tuples so it cannot have null value

Primary key attributes of each **relation R in Schema S** cannot have null values in **any tuple of State r(R).**

Integrity constraints ensure accuracy and consistency of data in a relational database.

Foreign Key Constraint



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

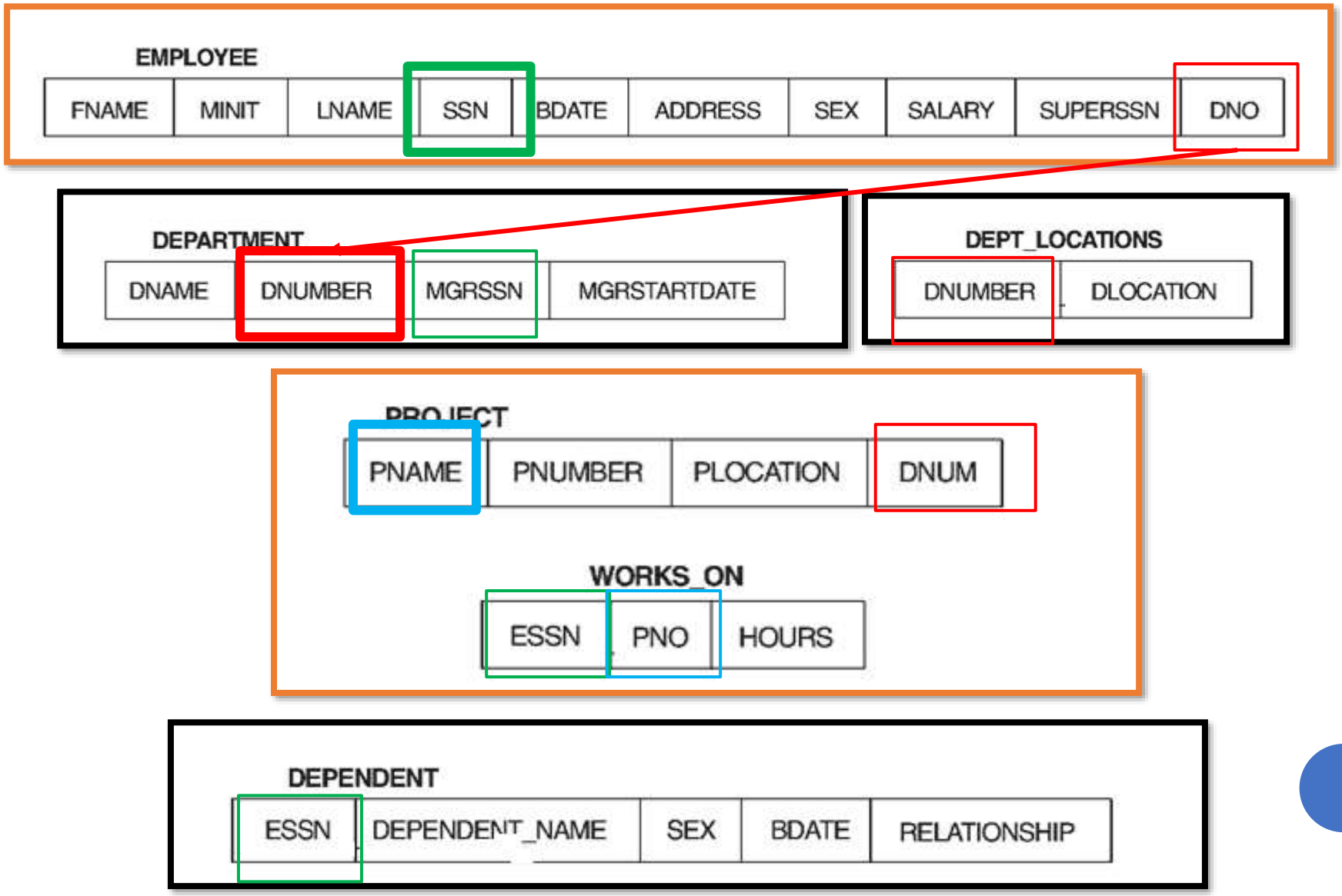
DEPARTMENT

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

Foreign Key Constraint



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

EMPLOYEE

Fname	Minit	Lname	Ssn	Super_ssn	Dno
John	B	Smith	123456789	333445555	5
Franklin	T	Wong	333445555	888665555	5
Alicia	J	Zelaya	999887777	987654321	4
Jennifer	S	Wallace	987654321	888665555	4
Ramesh	K	Narayan	666884444	333445555	5
Joyce	A	English	453453453	333445555	5
Ahmad	V	Jabbar	987987987	987654321	4
James	E	Borg	888665555	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

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

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

Referential Integrity

- ***A tuple in one relation that refers to another relation must refer to an existing tuple in that relation.***

- This is specified to maintain consistency among tuples in the two relations.

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
6	Houston

Foreign Key (FK)

A foreign key FK is a field in a table that matches the primary key column of another table.

- Attributes in FK in relation R_1 must has **same domain** as the attributes in PK of R_2
- **Value of FK** must be an **existing PK value** in R_2 **or Null**.

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

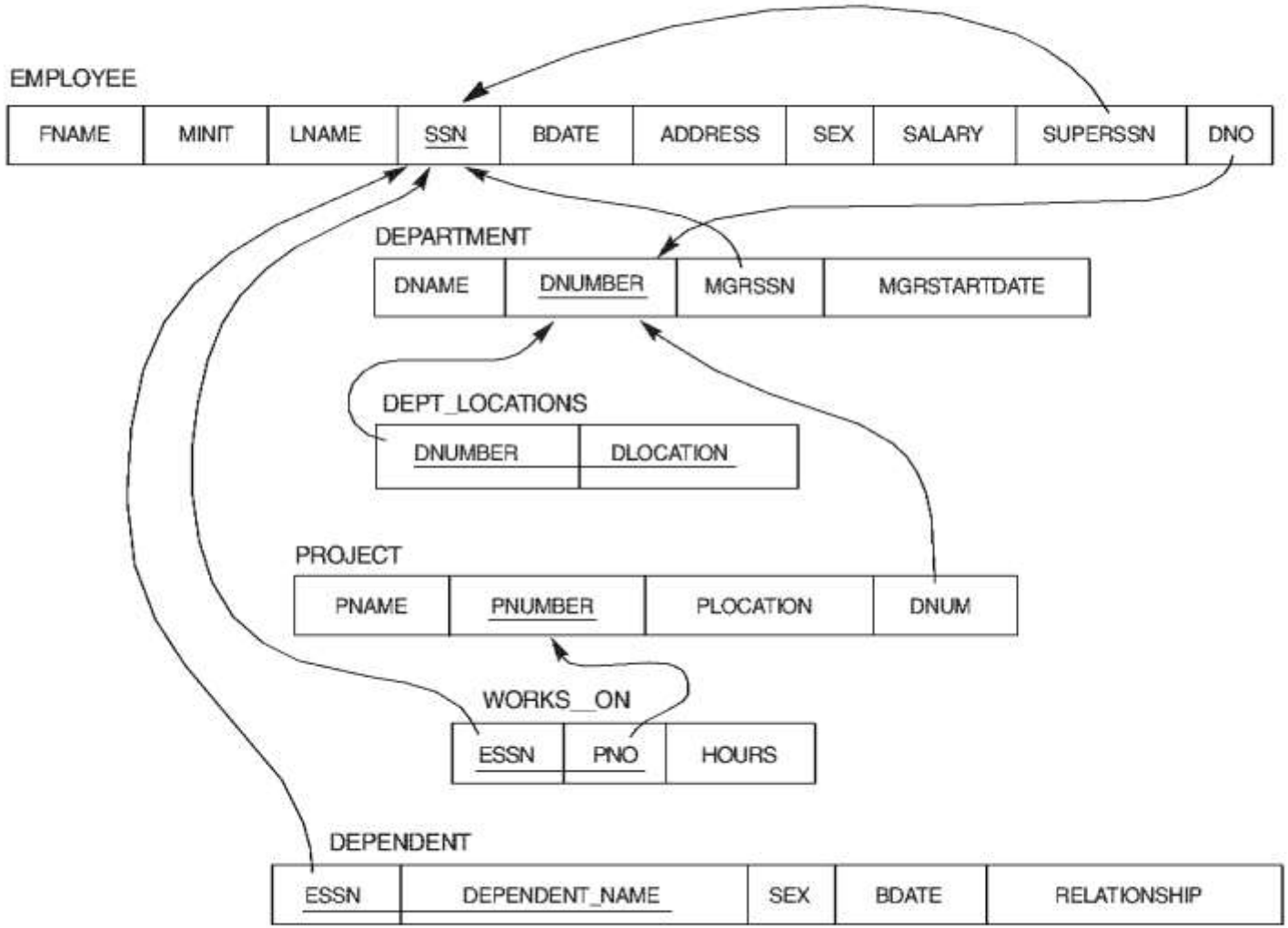
R2: Reference relation

DEPT_LOCATIONS

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

R1: Referencing relation

Figure Referential integrity constraints displayed on the COMPANY relational database schema diagram.



Foreign Key (FK)

- A foreign key can be a **primary key** or any of the candidate key in the referenced relation.

- FK requires that **uniqueness constraint** should hold for the column on which it is defined in referenced relation.

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

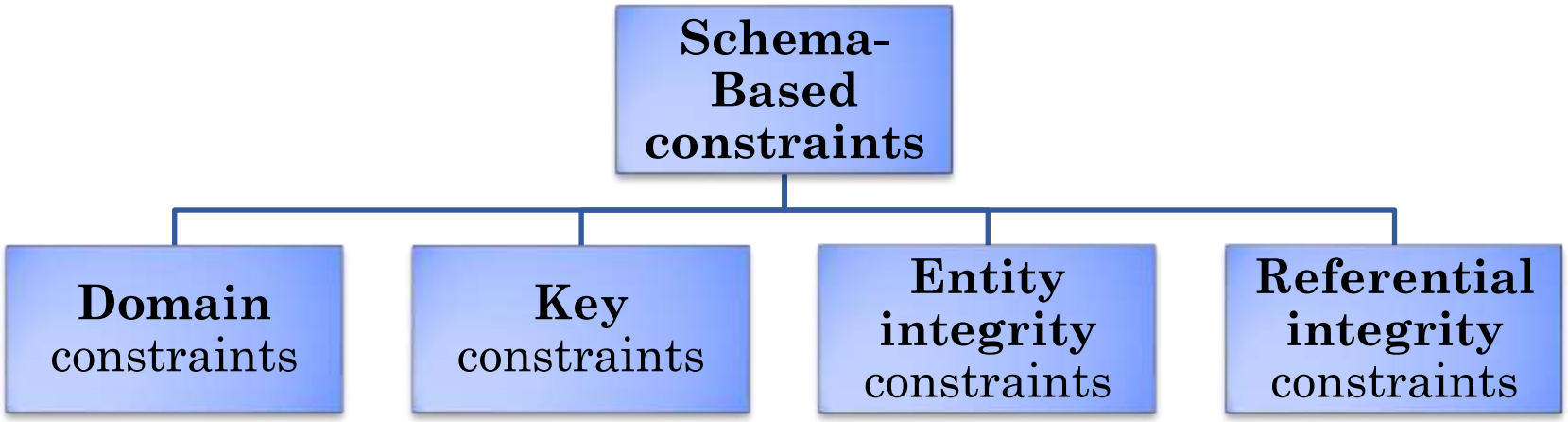
R2: Reference relation

DEPT_LOCATIONS

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

R1: Referencing relation

Schema-Based constraints



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

Relational Database Constraints

Model-based constraints or implicit constraints.

- These are inherent in data model, like no duplicate rows in table, domain is atomic.

Schema-based or explicit constraints.

- Can be expressed directly in the schema using DDL

Application based or semantic constraints or business rules.

- Can't be expressed directly in the schema
- Must be enforced by the application programs or **SQL triggers**

Semantic Integrity Constraints

- Based on application semantics and cannot be expressed by the model schema
- SQL allows triggers to specify some of these

- **Example**
  - Employee salary should not exceed supervisor salary

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

Semantic Integrity Constraints

- Based on application semantics and cannot be expressed by the model schema
- SQL triggers can specify these constraints

Example:

The max no of hours per employee for all projects he or she works on is 56 hrs per week

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