Database Systems

Lecture #05

Sang-Wook Kim
Hanyang University

Objectives

- To learn the relational data model
  - Concepts of a relational model
  - Notations of a relational model
  - Constraints of a relational model

Outline

- Relational Model Concepts
- Characteristics of Relations
- Relational Model Constraints
  - Key Constraints
  - Relational Database Schema
  - Integrity Constraints

Relational Model Concepts

- Database
  - Collection of *relations* (or *tables*)
- Relation
  - Set of *tuples* (or *rows*)

Relational Model Concepts

- Tuple
  - A fact that typically corresponds to a real-world *entity or relationship*
  - Consists of a set of *attributes*
- Attribute
  - Defines a property of each object corresponding to a *tuple*

Relational Model Concepts

STUDENT

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

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

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

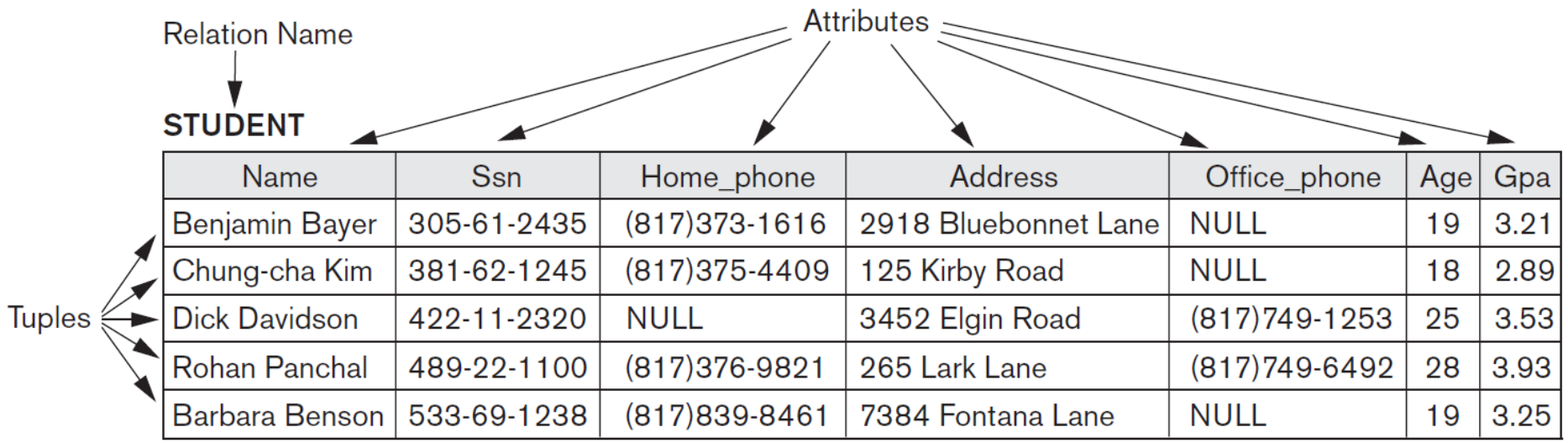
SECTION

Section_identifier	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

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

Relational Model Concepts



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

Relational Model Concepts

- Domain *D*
  - A set of *atomic* values
  - Example: domain of Age attribute
    - set of values between 16 ~ 80

Relational Model Concepts

- Attribute A_i
  - *Name of a role* played by some domain D in the relation schema R
  - dom(A_i): Domain of attribute A_i in relation schema R
  - Example: Employee_ages attribute in relation schema R
    - Has domain dom(Employee_ages)(=D)

Relational Model Concepts

- Relation schema R
  - Made up of a relation name R and a list of attributes, A_1, A_2, ..., A_n
  - Denoted by $R(A_1, A_2, ..., A_n)$
  - Example
    - STUDENT(Name, SSN, HomePhone, Address, OfficePhone, Age, GPA)

Relational Model Concepts

- Degree (or arity) of a relation
  - Number of attributes *n* of its relation schema
  - Example: degree of STUDENT relation
    - 7 (seven attributes)

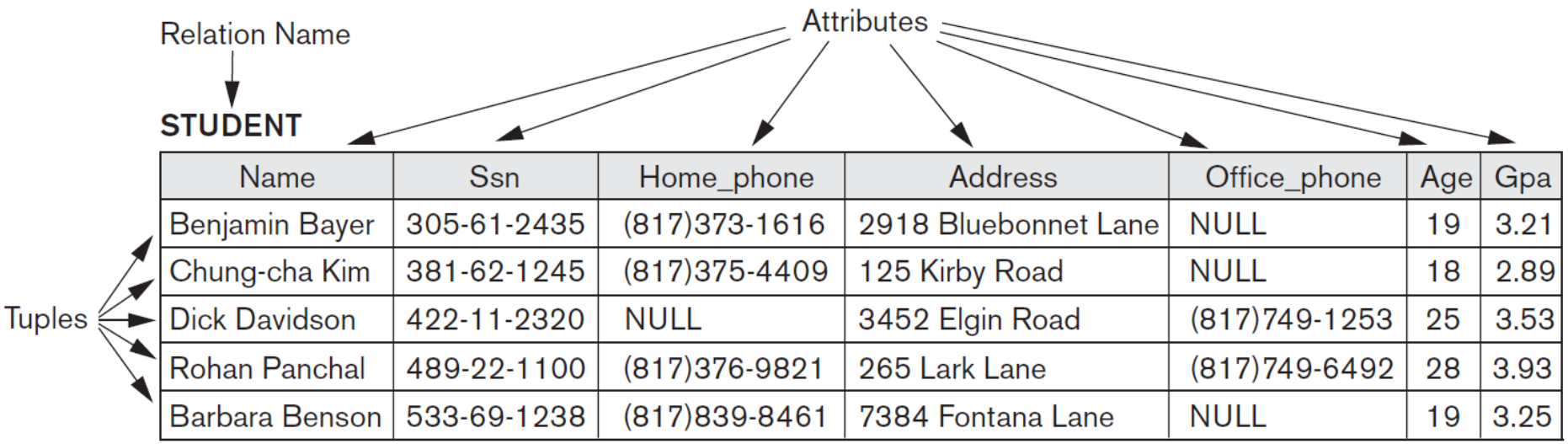
Relational Model Concepts

- Tuple t of relation R
  - Also called *n-tuple*
  - Ordered list of n values $t = <v_1, v_2, ..., v_n>$
  - Each value v_i, $1 \leq i \leq n$, is
    - an element of dom(A_i) or
    - a special *NULL* value

Relational Model Concepts

- Relation instance $r(R)$
  - Set of *tuples*
    - $r(R) = \{t_1, t_2, \ldots, t_m\}$
  - *Subset* of the *Cartesian product* of the domains that define R:
    - $r(R) \subseteq (\text{dom}(A_1) \times \text{dom}(A_2) \times \ldots \times \text{dom}(A_n))$

Relational Model Concepts



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

Characteristics of Relations

- Ordering of *tuples* in a relation
  - Relation defined as a *set* of tuples
  - Tuples have *no order* among them
    - Although tuples can be represented as a table

Characteristics of Relations

- Ordering of *attributes* within a tuple
  - *n*-tuple defined as a *ordered list of n values*
  - Order of attributes and values is not that important
    - As long as the correspondence between attributes and values maintained

Characteristics of Relations

- Values in tuples
  - Each value in a tuple is *atomic*
    - Composite and multivalued attributes not allowed
    - *First normal form* assumption
  - NULLs
    - Represent the values of attributes that may be *unknown* or may be *not applicable* to a tuple

Relational Model Notation

- $t[A_i]$
  - refers to the value v_i in t for attribute A_i
- $t[A_u, A_w, ..., A_z]$
  - refers to the sub-tuple of values $<v_u, v_w, ..., v_z>$ from t corresponding to the attributes specified in the list

Relational Model Constraints

- Constraints
  - *Restrictions* on the actual values in a relation schema
  - Every instance in the relation should follow them
  - Derived from the *rules* in the mini-world

Relational Model Constraints

- Important constraints in relational model

<integrity constraint>

  - Key constraint
  - Entity integrity constraint
  - Referential integrity constraint

domain constraint

Key Constraints

- *Superkey* of relational schema *R*
  - A set of attributes SK in *R*
  - No two distinct tuples in any relation instance *r*(*R*) of *R* can have the same value for SK
  - For any t_1 and t_2 ($t_1 \neq t_2$) in *r*(*R*)
    - $t_1[SK] \neq t_2[SK]$
  - Called *uniqueness constraint*

Key Constraints

- *Candidate key K*
  - Superkey of *R*
  - *Minimal* superkey
    - Removing any attribute *A* from *K*
    - Leaves a set of attributes *K* that *is not a superkey* of *R* any more
  - A relation schema may have more than one key
    - Candidate

Key Constraints

CAR

License_number	Engine_serial_number	Make	Model	Year
Texas ABC-739	A69352	Ford	Mustang	02
Florida TVP-347	B43696	Oldsmobile	Cutlass	05
New York MPO-22	X83554	Oldsmobile	Delta	01
California 432-TFY	C43742	Mercedes	190-D	99
California RSK-629	Y82935	Toyota	Camry	04
Texas RSK-629	U028365	Jaguar	XJS	04

Key Constraints

- Example: STUDENT relational schema
  - {SSN, Name, Age}: is a superkey / *not a candidate key*
  - {SSN}: is a superkey / is a candidate key

Key Constraints

- Primary key (PK) of the relation
  - Designated among ***candidate keys***
    - By a database designer
  - Underlined attribute
  - Example: DEPARTMENT relational schema
    - DepartmentID, DepartmentName

Key Constraints

- Key constraint
  - Specifies a relation must have candidate key attributes

Relational Database Schema

- Relational database schema
  - Set of *relation schemas* belonging to the same database
    - $S = \{R_1, R_2, ..., R_m\}$
  - Contains a set of *integrity constraints* IC

Relational Database Schema

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date

DEPT_LOCATIONS

Dnumber	Dlocation

PROJECT

Pname	Pnumber	Plocation	Dnum

WORKS_ON

Essn	Pno	Hours

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship

Relational Database Schema

- Relational database instance
  - A set of relation instances within the same database
    - $\{r_1, r_2, \ldots, r_m\}$
  - r_i is an instance of R_i which satisfies the integrity constraints specified in IC

Relational Database 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	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

Relational Database Schema

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

Entity Integrity Constraints

- Constraint for the attributes in PK of each relation schema *R*
- No primary key value can be NULL
- For every tuple *t* in *r*(*R*)
  - *t*[PK] ≠ NULL
- The primary key value is used to identify individual tuples in a relation

Referential Integrity Constraints

- Constraints specified *between two relations*
  - Key constraints and entity integrity constraints are specified on an individual relation
- Used to maintain the *consistency* among tuples in the two relations

Referential Integrity Constraints

- A tuple in one relation should refer to an *existing tuple* in that relation
- Example: DNO attribute in EMPLOYEE should refer to *existing value* for DNumber attribute in DEPARTMENT

Referential Integrity Constraints

- **Foreign key (FK)**
  - A set of attributes FK that satisfies the following rules for relation schema R_1 and R_2:
    - The attributes in FK in R_1 have the same domain(s) as the primary key attributes PK in R_2
    - FK in a tuple t_1 of the current instance $r_1(R_1)$ either has *a value of PK* for some tuple t_2 in the current instance $r_2(R_2)$ or is *NULL*

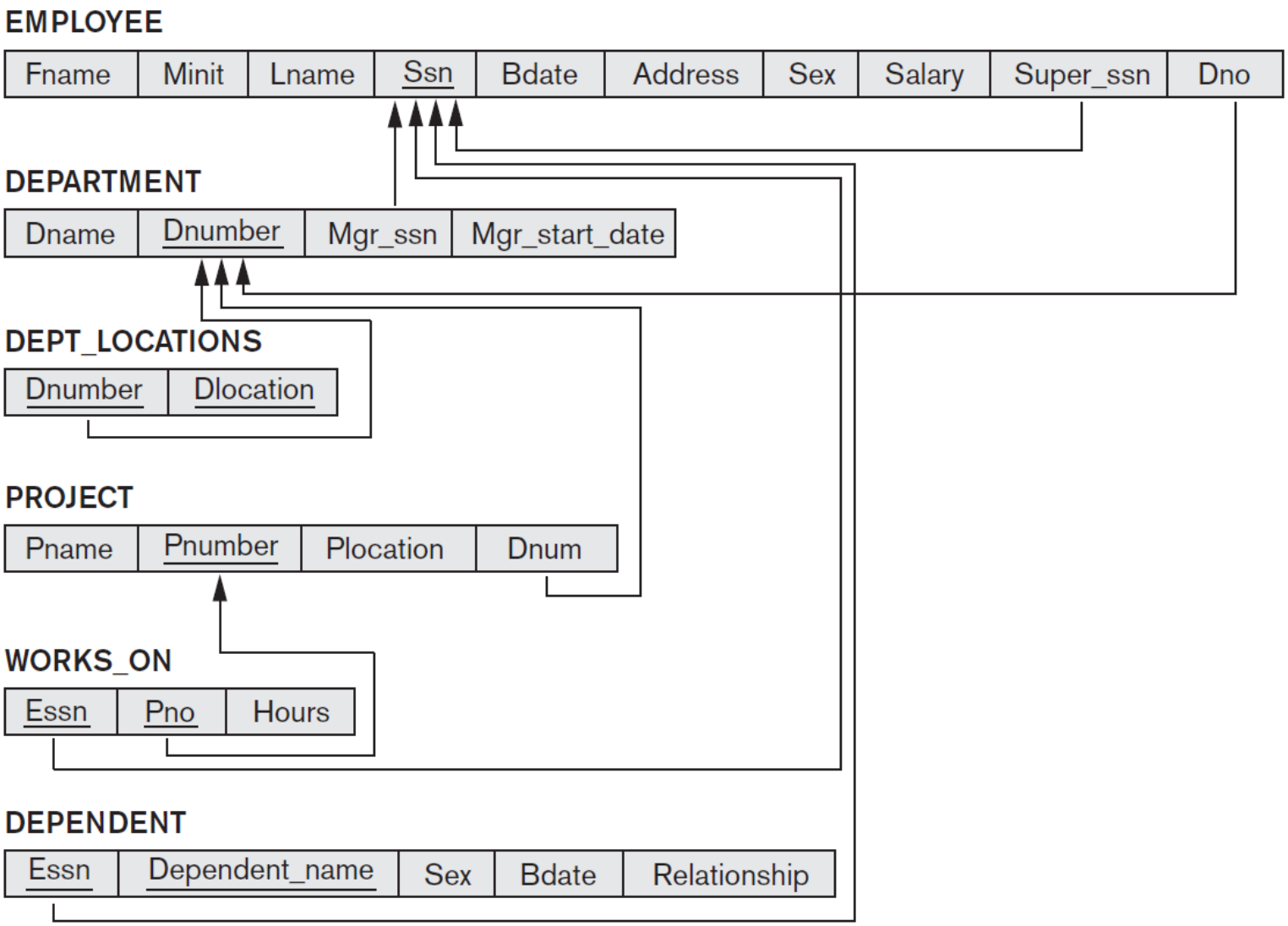
Referential Integrity Constraints

- **Foreign key (FK) (cont'd)**
  - t_1[FK] = t_2[PK] (or NULL)
  - *'t_1 **refers** to t_2'*

Referential Integrity Constraints

- Referential integrity constraint
  - Constraints specifying the foreign key
  - Represented as an *arrow from FK of R_1 to PK of R_2*

Referential Integrity Constraints



Summary

- Relational Model Concepts
  - Relations, tuples, attributes, domains, relational schemas, and relational instances
- Characteristics of Relations
- Relational Model Constraints
  - Key Constraints
  - Integrity constraints
    - Entity integrity constraints
    - Referential integrity constraints

References

1. Codd, Edgar F. "A relational model of data for large shared data banks." *Communications of the ACM* 13.6 (1970): 377-387.
2. Codd, Edgar F. "A data base sublanguage founded on the relational calculus." *Proceedings of the 1971 ACM SIGFIDET (now SIGMOD) Workshop on Data Description, Access and Control*. ACM, 1971.
3. Codd, Edgar F. *Relational completeness of data base sublanguages*. IBM Corporation, 1972.
4. Codd, Edgar F. "Recent Investigations in Relational Data Base Systems." *IFIP congress*. Vol. 74. 1974.
5. Codd, Edgar F. "Extending the database relational model to capture more meaning." *ACM Transactions on Database Systems (TODS)* 4.4 (1979): 397-434.

References

6. Childs, David L. *Feasibility of a set-theoretic data structure. A general structure based on a reconstituted definition of relation*. No. TR-6. MICHIGAN UNIV ANN ARBOR, 1968.
7. Codd, Edgar F. *The relational model for database management: version 2*. Addison-Wesley Longman Publishing Co., Inc., 1990.
8. Date, Chris J. *The Database Relational Model: A Retrospective Review and Analysis: A Historical Account and Assessment of EF Codd's Contribution to the Field of Database Technology*. Addison-Wesley, 2001.
9. Todd, Stephen JP. "The Peterlee relational test vehicle—a system overview."*IBM Systems Journal* 15.4 (1976): 285-308.

References

10. Schmid, Hans Albrecht, and J. Richard Swenson. "On the semantics of the relational data model." *Proceedings of the 1975 ACM SIGMOD international conference on Management of data*. ACM, 1975.
11. Chen, Peter Pin-Shan. "The entity-relationship model—toward a unified view of data." *ACM Transactions on Database Systems (TODS)* 1.1 (1976): 9-36.
12. Maier, David. *The theory of relational databases*. Vol. 11. Rockville: Computer science press, 1983.
13. Atzeni, Paolo, and Valeria De Antonellis. *Relational database theory*. Benjamin-Cummings Publishing Co., Inc., 1993.

Have a nice day!