Database Management Systems (DBMS)

Lec 2: Relational model of data

Ramesh K. Jallu

IIIT Raichur

Date:22/01/21

Recap

- Basic definitions in DBMS: DB system = data + DBMS
- Components in a DB system: H/w, s/w, data, users, and procedures
- Goals of a database system: Efficiency and conveniency
- File-processing system: redundancy and inconsistency, difficulty in accessing data, data isolation, intigrety, atomicity, security, and congruent access problems
- Advantages of using a DBMS approach: flexibility, up-to-date, etc.

Characteristics of a database

1. Self-describing nature of a database system
2. Insulation between program and data, and data abstraction
3. Support of multiple views of the data
4. Sharing of data and multiuser transaction processing

Today's plan

- Categories of data models
- Relation schema and its characteristics
- Relational model constraints: Keys

Data models

- A collection of tools for describing data, data relationships, data semantics, and data constraints

- Categories:

1. Relational model
2. Entity-Relationship model
3. Semi-structured model: flexible format for data exchange
4. Object-based model

1. Relational model

- Proposed by Ted Codd of IBM Research in 1970
- Commercial implementation took place in early 80's, SQL/DS
- It is the primary data model for commercial data-processing appl.
- It retains the primary position over its half-century of existence
- Uses concept of relation in mathematics

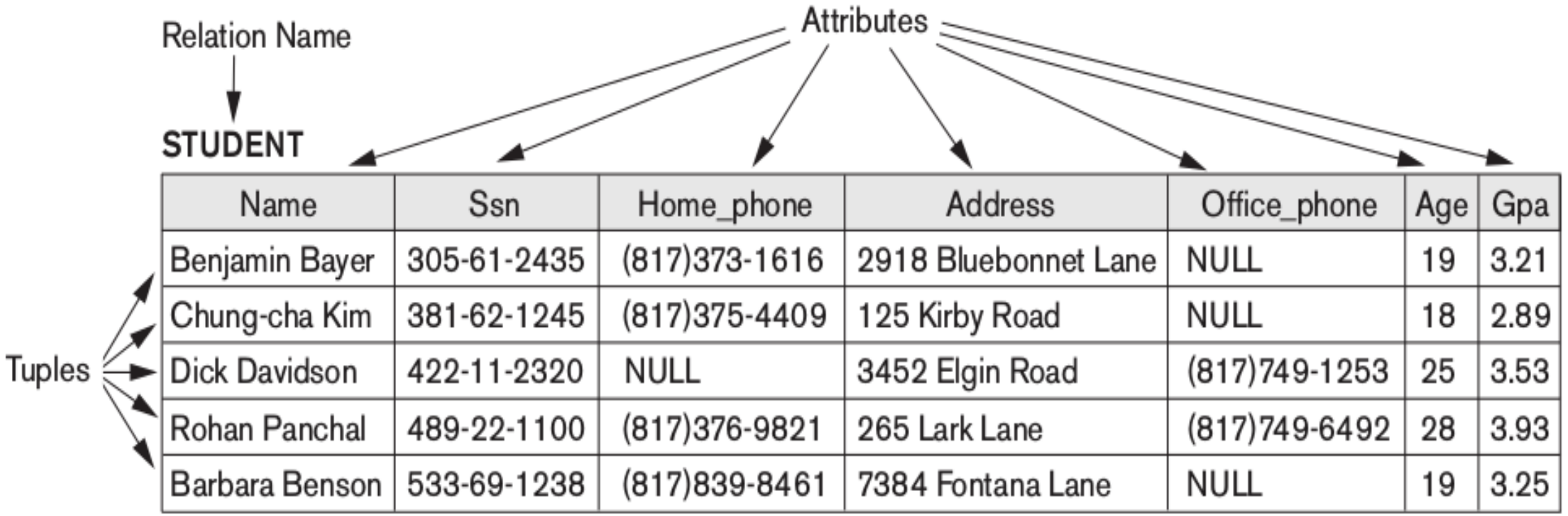
Relational model concepts

- This model represents the database as a collection of ***relations***
- Each relation resembles a ***table*** of values and assigned a unique name
- Each row in the table represents a collection of related data values
- The names of the table and columns in the table are used or chosen to interpret the meaning of the values in each row

Terminologies

- Tuple
- Attribute
- Domain of an attribute
- Atomic domain
- NULL value

An example



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

Another example

Instructor

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

Instructor

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

Atomic domain

Instructor

ID	*name*	*dept_name*	*salary*	*phone_number*
10101	Srinivasan	Comp. Sci.	65000	8729244589,9394029483
12121	Wu	Finance	90000	
15151	Mozart	Music	40000	
22222	Einstein	Physics	95000	
32343	El Said	History	60000	NULL
33456	Gold	Physics	87000	
45565	Katz	Comp. Sci.	75000	
58583	Califieri	History	62000	
76543	Singh	Finance	80000	865-8746-0984
76766	Crick	Biology	72000	
83821	Brandt	Comp. Sci.	92000	NULL
98345	Kim	Elec. Eng.	80000	

Relational schema

- Relation schema defines the design and structure of a relation
  - For a relation ***R***, with attributes ***A1,A2,…,An***, the relation schema is denoted by ***R(A1,A2,…,An)***
  - STUDENT(Name, Ssn, Home_phone, Address, Office_phone, Age, Gpa)
  - STUDENT(Name: string, Ssn: string, Home_phone: string, Address: string, Office_phone: string, Age: integer, Gpa: real)
- The ***degree*** or ***arity*** of relation is the no. of attributes in its relation

Relational schema (cont.)

- Domain of an attribute A_i is denoted by ***dom(A_i)***
  - A relation state ***r***, ***r(R)***, of a relation schema ***$R(A_1,A_2,\ldots,A_n)$*** is a set of n-tuples $r=\{t_1,t_2,\ldots,t_m\}$
  - Each tuple t_i is an ordered list of n values $t_i=<v_1,v_2,\ldots,v_n>$, where each v_j belongs to $dom(A_j)$
  - $r(R) \subseteq dom(A_1)\ X\ dom(A_2)\ X\ \ldots\ X\ dom(A_n)$

Characteristics and constraints

- Characteristics of a relation
  a. Ordering of tuples in a relation
  b. Ordering of values in a tuple
  c. Values and NULLs in the tuples

- Constraints in a relational model
  - Domain constraints
  - Key constraints: $t_i \neq t_j$ for $i \neq j$
  - NULL constraints

Instructor

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

Instructor (ID, name, dept_name, salary)

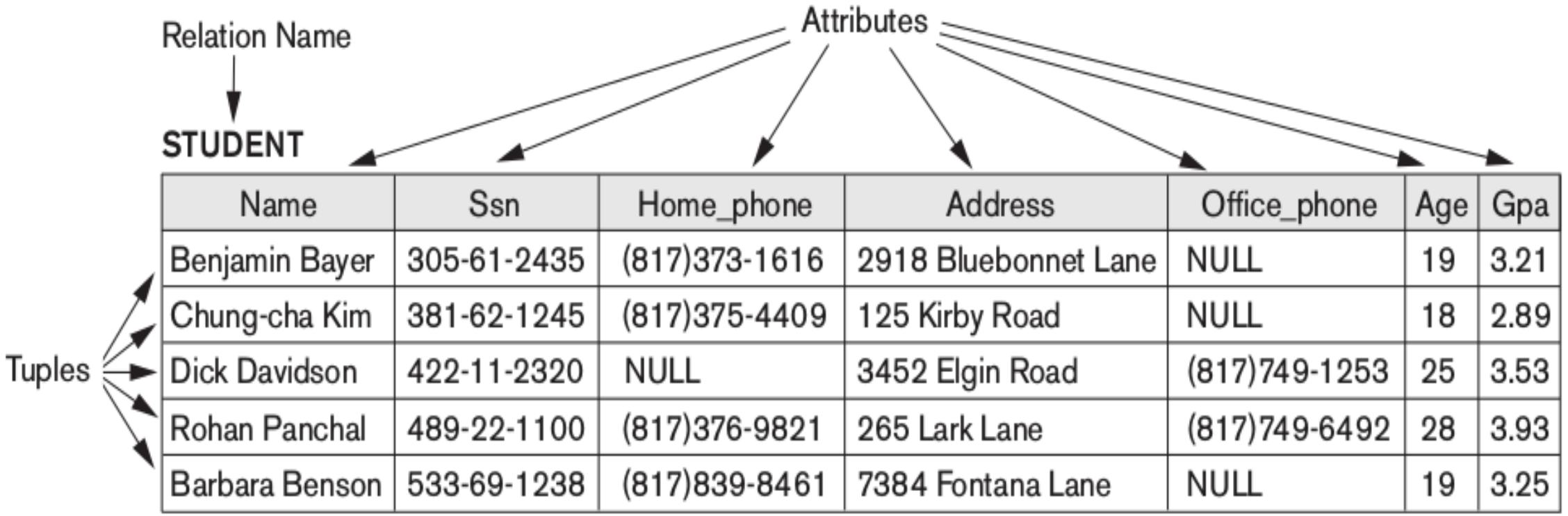
Keys

- Keys are used to distinguish tuples in a given relation
  - **Superkey**: A subset ***K*** of attributes that ***uniquely*** identifies a tuple in the relation
  - If t_i and t_j are any two distinct tuples in a relation, then $t_i[K] \neq t_j[K]$
  - Any relation contains a trivial superkey

Instructor

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

An other example



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

Keys (cont.)

- If ***K*** is a superkey, then so is any superset of ***K***
- We are interested in ***minimal*** superkey (a.k.a. Key)
- A relation schema may have more than one key. Such minimal superkeys are called ***candidate keys***
- One of the candidate keys is designated as ***primary key*** and other candidate keys are designated as ***unique keys***

Keys (cont.)

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

CAR(License_number, Engine_serial_number, Make, Model, Year)

CAR(License_number, Engine_serial_number, Make, Model, Year)

Primary key constraints

- It is customary to list the primary key attributes of a relation schema before the other attributes
- The primary key should be chosen such that its attribute values are never, or are very rarely, changed
  - E.g., address should not be chosen
- Entity integrity constraint: primary key value cannot be NULL

Relational databases

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

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

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

Relational database schema

- For a given relational database with relations $R_1, R_2, \ldots, R_m$
  - A relational database schema S is a set of relation schemas

  $S = \{R_1, R_2, \ldots, R_m\}$ and a set of integrity constraints

  - e.g., COMPANY = {EMPLOYEE, DEPARTMENT, DEPT_LOCATIONS, PROJECT, WORKS_ON, DEPENDENT}

  - **Referential integrity constraint**: used to maintain consistency among tuples in two relations

Foreign key

- Let R1 and R2 be two relations. A set of attributes K in R1 is a ***foreign key*** of R1
  - The attributes in K have the same domain(s) as the primary key attributes of R2
  - A value of K in a tuple ti of R1 either occurs as a value of primary key in some tuple of tj of R2 or is NULL
  - ti[K] = tj[primary key]
- R1 is called the referencing relation and R2 is called the referenced relation

What we learned today?

- Relational data model
- Constraints in the model

Thank you!