CSIT115 Data Management and Security

The Relational Model of Data

Dr Janusz R. Getta

School of Computing and Information Technology -
University of Wollongong

The Relational Model of Data

Outline

Basic Concepts

Data model ? What is it ?

A data model provides an abstract view of data that can be used for data definition, data manipulation, data retrieval, and data administration

Accordingly to Wikipedia (https://en.wikipedia.org/wiki/Data) a data model organizes data elements and standardizes how the data elements relate to one another

Because a data model provides an abstract view it is also commonly called as a view of data

In the past we talked about the following views of data:

- Sector, track, cylinder
- Sequence of data blocks
- Record, file, file system
- Two dimensional tables (tabular view), Hierarchies (tree view), Networks (graph view)
- Classes of objects, associations, attributes

Basic Concepts

In 1970 E.F. Codd from IBM Corp. defined a model of data based on a tabular view and called it as Relational Model of Data or simply Relational Model

At the moment (early 2019) ~95% of all database systems is based on Relational Model of Data

However, it is important to say that in the past and now Relational Model of Data had and still has few serious contenders like Object-Oriented Model, Object-Relational Model, XML Data Model, and recently JSON Data Model

Basic Concepts

What view of data provides Relational Model of Data ?

- The model provide a tabular view of data
- A relational table consists of a header and theoretically unlimited number of rows
- A header consists of a sequence of attribute names
- A row consists of a sequence of of values of attributes
- A vertical sequence of attribute name followed by attribute values is called as a column
- A header is also called as a relational schema
- A set of all values of an attribute is called as a domain of an attribute
- A database is a set of relational tables

Basic Concepts

A sample relational table:

anum	fname	lname	dob	city	state	phone
1	Harry	Potter	1980-12-12	Perth	Western Australia	645278453
2	Johnny	Walker	1990-01-1	Geelong	Victoria	63569784
3	Mary	Poppins	1950-01-01	Melbourne	Victoria	62389541
4	Michael	Collins	1960-05-25	Brisbane	Queensland	63336666
5	Margaret	Finch	1953-12-07	Sydney	New South Wales	64573489
6	Claudia	Kowalewski	1959-05-03	Hobart	Tasmania	64577744
7	James	Bond	1960-01-01	Perth	Western Australia	645 78434
8	Stephen	Staunton	1977-10-23	Freemantle	Western Australia	NULL
9	Joseph	Staunton	1977-10-23	Newcastle	New South Wales	623778453
10	John	Spiderman	1990-06-21	Sydney	New South Wales	24256789
11	George	TheFirst	1991-10-12	Melbourne	Victoria	NULL
12	Homer	Simpson	1957-05-24	Adelaide	South Australia	61369876
13	Neil	Superman	1960-07-20	Perth	Western Australia	45672345
14	Ivan	TheTerrible	1969-05-11	Brisbane	Queensland	123567898
15	Penelope	Princess	1977-10-23	Hobart	Tasmania	40076711
16	Zhi Chao	Zhong	1971-07-21	Horsley	New South Wales	86150189
17	Richard	TheLionheart	1981-06-02	Waga Waga	New South Wales	61234567
18	Sherlock	Holmes	1935-06-13	Bundaberg	Queensland	46676601
19	Robin	Hood	1951-08-21	Horsley	New South Wales	86150329
20	Janusz	Getta	1953-10-03	Horsley	New South Wales	12345678

Basic Concepts

Why a relational table is called as a "relational" ?

This is because of the following original E.F. Codd's definition of a relational table:

- Let A_1, A_2, ... A_n be the names of attributes
- Let dom(A_1), dom(A_2), ... dom(A_n) be the domains of the attributes, i.e. the sets of values of each attribute A_1, A_2, ... A_n
- A relational table is defined as a subset of Cartesian product dom(A_1) x dom(A_2) x ... x dom(A_n)

In mathematics a subset of Cartesian product is called as a relation

This is why a relational table is called as a "relational"

The original definition of a relational table is not correct because two tables with the different orders of columns contain the same information howeever two relations with different order of domains are different

The Relational Model of Data

Outline

Relational Table

A correct definition of relational tables is the following:

- Let A_1, A_2, ... A_n be the names of attributes
- Let dom(A_1), dom(A_2), ... dom(A_n) be the domains of attributes, i.e. the sets of values of each attribute A_1, A_2, ... A_n
- A row r is a full mapping r: {A_1, A_2, ...A_n} -> dom(A_1) U dom(A_2) U ... dom(A_n) such that for all A in {A_1, A_2, ... A_n}, r(A) is in dom(A)
- A relational table is defined as a set of rows

Now, an order of columns is immaterial, but ... a name "relational" is not well justified ;)

The Relational Model of Data

Outline

Principles of Relational Model

A relational table that has no multivalued attributes and composite attributes is in the first normal form (1NF)

For example, a relational table below is NOT in 1NF, sometimes we say that such table is in 0NF

e#	name	car used
950001	Peter	Toyota, PKR234 Ford, WER545
932345	Paul	Honda, RTQ456
960020	Joan	Holden, KLR197 Holden, KLR567

Are there any higher normal forms like for example 2NF, 3NF, ... ?

YES ! However, we shall not discuss this topic now !

Is 0NF completely useless ?

NO ! However, we shall not discuss this topic in this subject !

Principles of Relational Model

Access to the rows by the contents rule:

- We can only retrieve rows by their contents

It is NOT allowed to say: give me the second row from the following table:

APPLICANT Table

anum	fname	lname	dob	city	state
1	Harry	Potter	1980-12-12	Perth	Western Australia
2	Johnny	Walker	1990-01-13	Geelong	Victoria
3	Mary	Poppins	1950-01-01	Melbourne	Victoria
4	Michael	Collins	1960-05-25	Brisbane	Queensland
5	Margaret	Finch	1953-12-07	Sydney	New South Wales
6	Claudia	Kowalewski	1959-05-03	Hobart	Tasmania
7	James	Bond	1960-01-01	Perth	Western Australia
8	Stephen	Staunton	1977-10-23	Freemantle	Western Australia
9	Joseph	Staunton	1977-10-23	Newcastle	New South Wales
10	John	Spiderman	1990-06-21	Sydney	New South Wales

We have to say: give me a row such that `anum = 2` or such that

```
fname = 'Johnny' and lname = 'Walker'
```

Principles of Relational Model

Unique rows rule:

- Relational table cannot contain two identical rows

This rule is violated by all commercial Database Management Systems !

The Relational Model of Data

Outline

Keys

Let R(A_1, A_2, ... A_n) be a relational table with a relational schema (header) {A_1, A_2, ... A_n}

A key for a table R is a set of attributes K = {A_{k1}, A_{k2}, ... A_{km}} such that:

- (1) K is included in {A_1, A_2, ... A_n}, i.e. K is a subset of the schema
- (2) for any two rows v, w in R(A_1, A_2, ... A_n) their k-values must be different, i.e. v[k] <> w[k]
- (3) no proper subset of K satisfies a property (2) above

A key that does not satisfy a condition (3) is called as superkey

A key that satisfies the conditions (1) and (2) and (3) is called as minimal key

Keys

Examples:

- A set of attributes `{snum}` is a minimal key in a relational schema `STUDENT={snum, first-name, last-name, date-of-birth}`
- A set of attributes `{snum, last-name}` is a superkey in a relational schema `STUDENT={snum, first-name, last-name, date-of-birth}`
- A set of attributes `{snum, code, enrolment-date, enrolment-time}` is a minimal key in a relational schema `ENROLMENT={snum, code, enrolment-date, enrolment-time}`
- A set of attributes `{bldg#, room#}` is a minimal key in a relational schema `ROOM={bldg#, room#, area}`
- A set of attributes `{p#, manufacturer, price}` is a superkey in a relational schema `PART={p#, name, price, manufacturer}`
- A set of attributes `{p#, manufacturer}` is a superkey in a relational schema `PART={p#, name, price, manufacturer}`
- A set of attributes `{p#}` is a minimal key in a relational schema `PART={p# name, price, manufacturer}`

Keys

More examples:

- A set of attributes {pnum, first-name,last-name, dob, team} is a superkey in a relational schema PLAYER={pnum, first-name, last-name, dob, team}
- A set of attributes {pnum, first-name,last-name, dob} is a superkey in a relational schema PLAYER={pnum, first-name, last-name, dob, team}
- A set of attributes {first-name,last-name, dob} is a minimal key in a relational schema PLAYER={pnum, first-name, last-name, dob, team}
- A set of attributes {supplier-num, part-num, delivery-date, delivery-address} is a minimal key in a relational schema SHIPMENT={supplier-num, part-num, delivery-date, delivery-address}

Keys

All minimal keys valid in a relational schema are also called as candidate keys

A primary key is one of the candidate keys arbitrarily chosen by a database designer to uniquely identify the rows in a relational table

Examples:

- A set of attributes `{snum}` and a set of attributes `{first-name, last-name, date-of-birth}` are the candidate keys in a relational schema `STUDENT={snum, first-name, last-name, date-of-birth}`
- A candidate key `{snum}` can be selected by a database designer as a primary key
- It is also possible that a candidate key `{first-name, last-name, date-of-birth}` can be selected by a database designer as a primary key

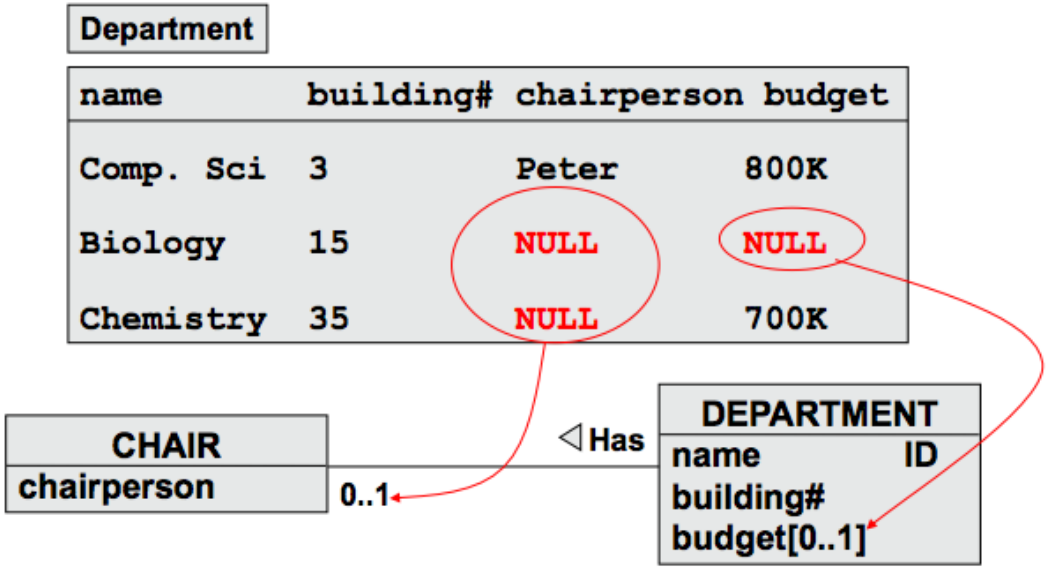
In the future a relational schema $R = \{A_1, A_2, \ldots A_n\}$ will be denoted by $R(A_1, A_2, \ldots A_n$ and any sort of key $\{A_{i1}, A_{i2}, \ldots A_{im}\}$ included in R will be denoted by $(A_{i1}, A_{i2}, \ldots A_{im})$

The Relational Model of Data

Outline

NULL

A NULL constraint says that an attribute in a relational table may have no values at all



With an exception saying that no column belonging to a primary key or candidate key is allowed to take on NULL for any row (it is also called as Entity Integrity constraint)

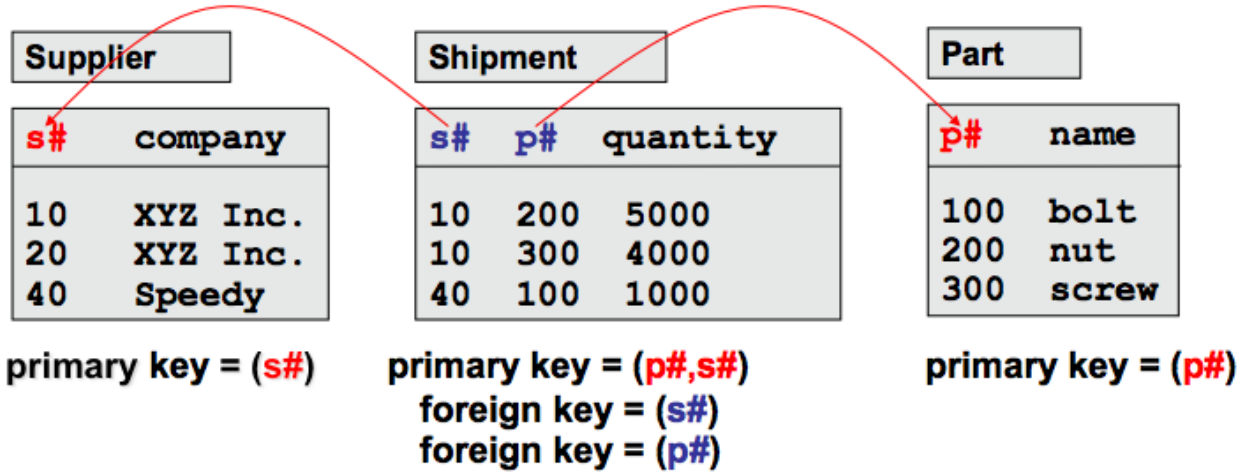
All commercial Database Management Systems allow NULL for candidate keys

The Relational Model of Data

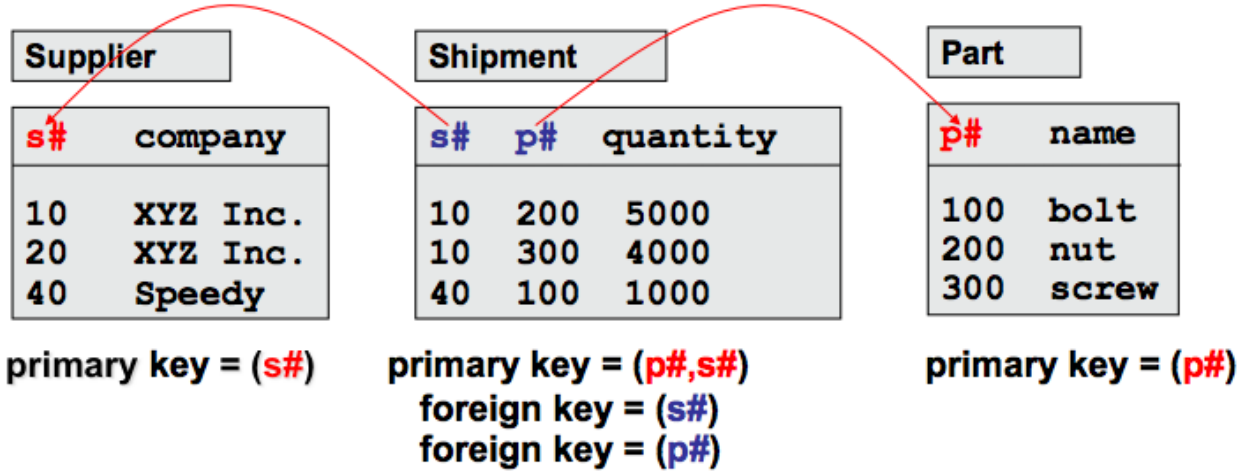
Outline

Referential Integrity Constraints

A set of attributes F in a relational schema R is a called as a foreign key if the combination of values of attributes in F in any row is required to either contain `NULL`s or else to match the value combination of a set of columns P representing a candidate or primary key in some other relational schema S



Referential Integrity Constraints



Referential integrity rule is in force if the columns of foreign key in any relational table either:

- (1) have NULLs in at least one column that allows NULLs
- (2) have no NULLs and combination of all its values is equal to the combination of primary key values in the other relational table

Referential Integrity Constraints

Example of referential integrity constraint:

- A relational schema `BUILDING(bldg#, floor#, name)` has a primary key `(bldg#)`
- A relational schema `ROOM(bldg#, room#, area}` has a primary key `(bldg#, room#)`
- Then a set of attributes `(bldg#)` included in a schema `ROOM` is a foreign key that references a primary key `(bldg#)` in a schema `BUILDING`

Another example of referential integrity constraint:

- A relational schema `STUDENT(s#, first-name, last-name, dob)` has a primary key `(s#)`
- A relational schema `SUBJECT(code, title, credits)` has a primary key `(code)`
- Then a relational schema `ENROLMENT(s#, code, edate)` has a foregin key `(s#)` referencing primary key `{s#}` in a schema `STUDENT` and ...
- ... a relational schema `SUBJECT=(code, title, credits)` has a foreign key `(code)` referencing primary key `(code)` in a schema `SUBJECT`

Referential Integrity Constraints

Yet another example of referential integrity constraint:

- A relational schema `ROOM=(bldg#, room#, area)` has a primary key `(bldg#, room#)`
- A relational schema `LECTURER=(emp#, first-name, last-name, bldg#, room#)` has a primary key `(emp#)`
- Then a set of attributes `(bldg#, room#)` included in a relational schema `LECTURER` is a foreign key that references a primary key `(bldg#, room#)` in a schema `ROOM`

The Relational Model of Data

Outline

Domain Constraints

A domain constraint is a condition imposed on the values of an attribute `A` that determines the values of `dom(A)`, i.e. a domain of attribute `A`.

Examples:

- An attribute `student-number` is a sequence of 7 digits
- An attribute `date-of-birth` cannot have a value greater then todays date
- An attribute `salary` is a positive real number
- A value of an attribute `gender` can be either 'female' or 'male'
- A value of an attribute `credits` can be either 6 or 12
- A value of an attribute `first-name` is a string of letters an blanks that starts from a capital letter

The Relational Model of Data

Outline

Summary

A database is a collection of relational tables

A relational table consists of rows (tuples) and columns (attributes)

All attributes have atomic values

Each attribute has a domain, i.e. a set of acceptable values

A row represents a relationship among a set of attributes

A relational table is a subset of Cartesian product of attribute domains

An attribute may have no value (`NULL`)

A relational table implements either a class of objects or an association

All identifiers in a conceptual schema are implemented as the keys in the relational tables

Summary

A "tourist guide" through a "land of keys"

- Minimal key => the smallest key
- Superkey => minimal key + other attribute(s)
- Candidate key => any minimal key
- Primary key => one of candidate keys
- Foreign key => an attribute or set of attributes referencing a primary keyor a candidate key in another or the same relational table

References

T. Connoly, C. Begg, Database Systems, A Practical Approach to Design, Implementation, and Management, Chapter 4 The Relational Model, Pearson Education Ltd, 2015