Databases

2019-05-06 22:53
1a0f14a

Definitions and concepts

A **database** supports storage and retrieval of **structured** data. The data are organized into a set of **tables,** and those tables can be **joined** together using keys. In each table, the **rows** (also known as records) each represent a single member of that entity (examples: a movie, a person, an address, an invoice, a customer account). The **columns** (also known as fields or attributes) represent one piece of data for that entity, with a particular type. A column can be a string of text, a number, a date/time, an IP address, etc.

Example:

*ID	FirstName	LastName	PhoneNum	BirthDate
1	Alice	Smith	555-1234	1995-04-07
2	Bob	Chen	555-2352	NULL
3	Charlie	Jones	555-1919	1990-12-15
4	Charlie	Jones	555-8822	NULL
5	Doug	Drummond	555-8300	1986-07-19
6	Elsa	NULL	NULL	1992-08-30
7	Fred	Flintstone	555-9100	NULL

A **primary key** is an attribute that **uniquely identifies** each row in a table. That is, there can be no duplicates. Often the best thing to use as a primary key is an arbitrary integer. A good example of this is your student ID from LIU. It's possible, but a bit tricky, to user other, non-arbitrary data as a primary key. For example, if we were to use a customer's first and last name, there could be duplicates. So to make it a legitimate primary key, we'd have to sometimes add a number to ensure uniqueness. LIU does this with your logins (such as `jane.smith4`) and IMDB uses Roman numerals to distinguish people with the same name – such as Emma Watson (II) vs. Emma Watson (I). Other examples of integer primary keys are your passport number, driver's license number, social security number, and account numbers with other organizations.

A **foreign key** is an attribute that references the primary key of a **different table.** It is through the use of foreign keys that we can join together multiple tables in a database. The handout below illustrates this. The first two tables (Movie and Person) have integer primary keys in the "ID" column. Then third table references those primary keys. To answer some of the questions below, you have to look into all three tables.

Samples & references

- Handout with a sample database of artists, museums, and paintings

- These unlucky people have names that break computers
- `artists.db` – an SQLite database file

SQL statements for artist DB

```
-- #1
select max(born) < min(died) from artist;

-- #2
select painting.title from painting, museum where
painting.museum = museum.id and
museum.name = 'Yale Art Gallery';

-- #3
select museum.city from painting, museum, artist where
painting.museum = museum.id and
painting.artist = artist.id and
artist.last = 'Degas';

-- #4
select painting.title from painting, museum where
painting.museum = museum.id and
museum.city = 'New York'
order by painting.year limit 1;

-- #5
select distinct artist.first, artist.last
from painting, museum, artist where
painting.museum = museum.id and
painting.artist = artist.id and
museum.country = artist.country;
```