Tutorial: Advanced DB Design

> Designed by ZHU Yueming
>
> Update by ZHU Yueming in Mar. 30th 2022.
>
> Update by ZHU Yueming in April. 4th 2023.

Experimental objective

- Understand 3 normal form of database design
- Learn how to design a database according to the requirement document

1. Normal Form

1 NF

A relation is in first normal form if and only if the domain of each attribute contains only **atomic** (indivisible) values, and the value of each attribute contains only **a single value** from that domain.

Comparison 1:

The first one doesn't satisfy 1NF, because the attribute is not atomic. More specifically, the name can be split to English name and Chinese name, location can be split to longitude and latitude.

First:

station_id	name	location
1	Luohu(罗湖)	114.11833 , 22.53111
2	Guomao(国贸)	114.11889 , 22.54
3	Laojie(老街)	114.11639 , 22.54444
4	Grand Theater(大剧院)	114.10333 , 22.54472
5	Science Museum(科学馆)	114.08972 , 22.54333
6	Huaqiang Rd(华强路)	114.07889 , 22.54306
7	Gangxia(岗厦)	114.06306 , 22.53778
8	Convention and Exhibition Center Station(会展中心)	114.05472 , 22.5375
9	Shopping Park(购物公园)	114.05472 , 22.53444
10	Xiangmihu(香蜜湖)	114.034 , 22.5417

Improved:

station_id	english_name	chinese_name	longitude	latitude
1	Luohu	罗湖	114.11833	22.53111
2	Guomao	国贸	114.11889	22.54
3	Laojie	老街	114.11639	22.54444
4	Grand Theater	大剧院	114.10333	22.54472
5	Science Museum	科学馆	114.08972	22.54333
6	Huaqiang Rd	华强路	114.07889	22.54306
7	Gangxia	岗厦	114.06306	22.53778
8	Convention and Exhibition Cent…	会展中心	114.05472	22.5375
9	Shopping Park	购物公园	114.05472	22.53444
10	Xiangmihu	香蜜湖	114.034	22.5417

Comparison 2:

The first one doesn't satisfy 1NF, because the stations has more value of the domain, more specifically,

column station1 and station2 are similar, so that we should merge them into one column.

First:

line_id	line_color	station1	station2
1	Green	Bao'an Stadium	Pingzhou
2	Orange	Chiwan	Shekou Port
3	DeepSkyBlue	Danzhutou	*<null>*

Improved:

line_id	line_color	stations
1	Green	Bao'an Stadium
1	Green	Pingzhou
2	Orange	Chiwan
2	Orange	Shekou Port
3	DeepSkyBlue	Danzhutou

2 NF

 A relation satisfying 2NF must:

- be in 1NF
- not have any non-prime attribute that is dependent on any proper subset of any [candidate key](#) of the relation. A non-prime attribute of a relation is an attribute that is **not a part of** any candidate key of the relation.

The primary keys in the first table are `station_id` and `line_id`, but the first one is a awful design because the columns `line_color` and operator are not related to `station_id` (primary key) while `english_name`, `chinese_name` and `distric`t are not related to `line_id`, so that those attribute is a part of condidate key.
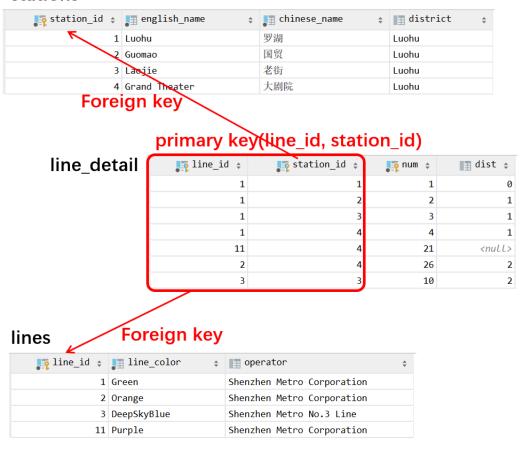
If we want to describe **many-to-many relationship** between two entities, we need to separate those two entities into two tables and then add a relation table to describe the relationship itself (as in the second case).

First:

station_id	english_name	chinese_name	district	line_id	line_color	operator
1	Luohu	罗湖	Luohu	1	Green	Shenzhen Metro Corporation
2	Guomao	国贸	Luohu	1	Green	Shenzhen Metro Corporation
3	Laojie	老街	Luohu	1	Green	Shenzhen Metro Corporation
4	Grand Theater	大剧院	Luohu	1	Green	Shenzhen Metro Corporation
4	Grand Theater	大剧院	Luohu	11	Purple	Shenzhen Metro Corporation
4	Grand Theater	大剧院	Luohu	2	Orange	Shenzhen Metro Corporation
3	Laojie	老街	Luohu	3	DeepSkyBlue	Shenzhen Metro No.3 Line

Improved:

stations

station_id	english_name	chinese_name	district
1	Luohu	罗湖	Luohu
2	Guomao	国贸	Luohu
3	Laojie	老街	Luohu
4	Grand Theater	大剧院	Luohu

Foreign key

primary key(line_id, station_id)

line_detail

line_id	station_id	num	dist
1	1	1	0
1	2	2	1
1	3	3	1
1	4	4	1
11	4	21	<null>
2	4	26	2
3	3	10	2

lines

Foreign key

line_id	line_color	operator
1	Green	Shenzhen Metro Corporation
2	Orange	Shenzhen Metro Corporation
3	DeepSkyBlue	Shenzhen Metro No.3 Line
11	Purple	Shenzhen Metro Corporation

3 NF

A relation satisfying 2NF must:

- be in 2NF
- all the attributes in a table are determined only by the candidate keys of that relation and not by any non-prime attributes.

We can see from the figure that the column `english_name` , `Chinese_name`, `district` is not related to the primary key `bus_line` directly. Those columns are more suitable for describing `station_id`. In this case, we'd better create two tables, the first one is `station` and the second is `bus_line`, and create a foreign key in bus_line table referencing the station table with station_id.
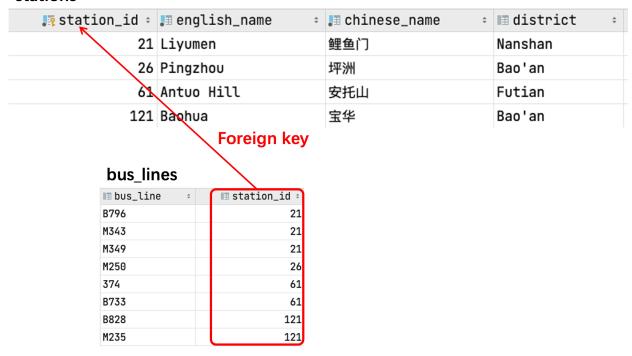
If we want to describe **one-to-many relationship**, e.g. A only has one B, but B could have more than one A, we can design a foreign key in A table referencing the primary key in B table.

First:

	bus_line	station_id	english_name	chinese_name	district
1	B796	21	Liyumen	鲤鱼门	Nanshan
2	M343	21	Liyumen	鲤鱼门	Nanshan
3	M349	21	Liyumen	鲤鱼门	Nanshan
4	M250	26	Pingzhou	坪洲	Bao'an
5	374	61	Antuo Hill	安托山	Futian
6	B733	61	Antuo Hill	安托山	Futian
7	B828	121	Baohua	宝华	Bao'an
8	M235	121	Baohua	宝华	Bao'an

Improved:

stations

station_id	english_name	chinese_name	district
21	Liyumen	鲤鱼门	Nanshan
26	Pingzhou	坪洲	Bao'an
61	Antuo Hill	安托山	Futian
121	Baohua	宝华	Bao'an

Foreign key

bus_lines

bus_line	station_id
B796	21
M343	21
M349	21
M250	26
374	61
B733	61
B828	121
M235	121

2. Comprehensive Exercise 12306

Background

As we know, the China's rail network is complex, more than 5000 station distribute in 657 citys and connected by 140000km railway. There are so many passengers take the train each day and processing these informations is a real challeng. You, a SUSTech student who want to design a **similarly** database about 12306 according to following **simplified** requirements.

Before Start

Although there are standards, database designing is still a highly subjective intellectual activity. As graders, we will set some rules for grading, but as long as your design can satisfy our requirements, you should be able to get the points. Since table name and field name is case-insensitive, it is recommended to use underline "_" to separate words, instead of using camel naming.(e.g. `Use my_great_table` , `some_informative_field` instead of `MyGreateTable` or `SomeInformativeField` .)

Example fields are only for reference. You may create new fields and combine/separate these fields to simplify your design, but the information needs to be remained at least

DETAILS ABOUT THE RELATIONSHIPS

Store data about rail lines and rail station in an organized and easy-to-maintain manner (see the keywords below). Following keywords only express what it is, how to create tables and fields are determined by your design.

- City and Station
  - A city have multiple rail stations
  - A city can have no stations
  - A station must be in a city
- Train and Station
  - A train has a train number and its own type with the type name (动车D、特快 T、高速 G、 GC-高铁、 null 其他)
  - A train has its depart station and arrival station
  - A train need provide the info of whether has air conditioner
  - There are several stations in one train
  - In a train, for each halfway station, we can find the last station, the next station, and the arrival time depart time for each station.

站序	站名	到站时间	出发时间	停留时间 ✖
01	沈阳南	----	06:50	----
02	沈阳	07:02	07:09	7分钟
03	阜新	07:56	07:58	2分钟
04	辽宁朝阳	08:32	08:35	3分钟
05	牛河梁	09:02	09:04	2分钟
06	承德南	09:35	09:37	2分钟
07	北京朝阳	10:28	10:28	----

G972次   沈阳南 -->   北京朝阳   高速   有空调

- Trip
  - A trip has the a train number, and departure station, arrival station with their time.
  - A trip has its own price, and the price of different seat type in one train is different
- Ticket
  - A concrete trip in concrete date. For example, "G74" in Feb. 28th or Feb. 29th.
  - Needs the datetime.
  - Need seat type and price... or other you think is necessary like the graph below

- Account
    - Serves as 12306 account, includes username and password.
    - Account must record it's ID card, the ID card number may have a 'x' in its last digit.
    - Needs phone number.
- Order
    - The Order should record account user, create date, order status, train num, the depart city, arrive city and price, etc.

Sample Answer

station

🔑 **id**		integer
name		varchar(30)
city		varchar(30)

train

🔑 **id**		integer
train_number		varchar(10)
type		varchar(2)
type_name		varchar(5)
air_conditioner		boolean

station_id:id

train_id:id

train_stations

🔑 **id**		integer
🔑 **station_id**		integer
🔑 **train_id**		integer
number		integer
depart_time		time
arrive_time		time

depart_station_id:id

trip

🔑 **id**		integer
depart_station_id		integer
arrival_station_id		integer
seat		varchar(5)
price		smallint

trip_id:id

ticket

🔑 **id**		integer
🔑 **trip_id**		integer
date		date
rest_ticket		integer

account

🔑 **id**		bigint
username		varchar(30)
password		varchar(18)
id_card		varchar(11)
name		varchar(6)
phone		varchar(11)

ticket_id:id

account_id:id

ticket_order

🔑 **id**		bigint
🔑 **account_id**		bigint
🔑 **ticket_id**		integer
create_date		date

Powered by yFiles