

数据库作业

环境：当前这个实例中9.145.42.250:3306，访问账号为job_r,密码为job_tmp_r,database 为:job

题目一

题目

现在新增一个主播业务，在MySQL上需要创建一个主播表tb_anchor，按照课程提供的规范，提供一个符合规范的建表语句(不需要实际执行，需要提供建表的SQL，规范参考：

<https://docs.qq.com/sheet/DQ3IHSXNoVXpDUIJz?tab=q9gfbn> 工作表2)。

anchor_id 主播id，取值范围：1-1000000

anchor_uid 主播的uid，取值范围：1-1000000

anchor_name 主播名称，取值范围：支持中文（500个字符）

nick_name 主播名称，取值范围：支持中文可以包含emoji符号（500个字符）

union_id 公会id，取值范围：1-1000000

union_name 公会简称，取值范围：支持中文（500个字符）

union_company_name 公会全称，取值范围：支持中文（500个字符）

contract_type 三方约类别 "不签约":0,"签约":1

idcard_no 主播身份证号

phone 主播手机号

operator 操作人，取值范围：支持中文（500个字符）

create_time 主播入驻时间

region 主播所在地域

address 主播地址，取值范围：支持中文可以包含emoji符号（500个字符）

email 主播地址，取值范围：（500个字符）

解答

```
1 CREATE TABLE tb_anchor(  
2   anchor_id INT PRIMARY KEY AUTO_INCREMENT COMMENT '主播ID',  
3   anchor_uid INT COMMENT '主播UID',  
4   name VARCHAR(500) COMMENT '主播名称',  
5   nick_name VARCHAR(500) COMMENT '主播名称(可包含emoji)',  
6   union_id INT COMMENT '工会ID',  
7   union_name VARCHAR(500) COMMENT '工会简称',  
8   union_company_name VARCHAR(500) COMMENT '工会全称',
```

```

9  contract_type INT(1) COMMENT '三方约类别 ""不签约"":0, ""签约"":1',
10 idcard_no VARCHAR(18) COMMENT '主播身份证号',
11 phone VARCHAR(20) COMMENT '主播手机号',
12 operator VARCHAR(500) COMMENT '操作人',
13 create_time date COMMENT '主播入驻时间',
14 region VARCHAR(50) COMMENT '主播所在地域',
15 address VARCHAR(500) COMMENT '主播地址',
16 email VARCHAR(500) COMMENT '主播邮箱'
17 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;

```

题目二

题目

通过题目1的数据表，业务提供了以下sql，需要对主播表添加合适的索引(不需要实际执行)，提供索引添加的SQL语句，并找出SQL不符合规范的地方,并改写，提供改写后的SQL:

```

1 select * from where tb_anchor where anchor_id = '10';
2 select anchor_id from tb_anchor where anchor_name = '非主流男孩';
3 select anchor_id from tb_anchor where union_id=10 and create_time >='2021-06-
4 select anchor_id,anchor_uid from tb_anchor where operator='管理员' ; -- 怎么
5 update tb_anchor set union_id=100,union_name='新人工会',contract_type=1 where
6 select region,count(region) from tb_anchor group by region having count(region)
7 select anchor_id,anchor_uid from user order by rand() limit 1;

```

解答

```

1 -- 索引添加:
2 alter tb_anchor add index name_index(anchor_name);
3 alter tb_anchor add index id_time_address(union_id,create_time, address);
4 -- 合规改写:
5 select * from tb_anchor where anchor_id = '10'; --第一句sql
6 update (select union_id,union_name,contract_type from tb_anchor where anchor_id=10)
7 select anchor_id,anchor_uid from tb_anchor limit rand(select count(*) from tb_anchor);

```

题目三

题目

由于导入了重复的数据 (union_name, union_id, operator_id, create_time 的数据应该唯一)，先需要导出一份 job 库 union_operator_info 的数据，格式为 CSV，编写导出命令，并计算导出文件的 md5，再编写去重的 SQL (不需要在目标库实际执行，保留自增 id 为最小值的数据)。

(需要提供数据导出命令，文件 MD5，去重 SQL)

解答

```
1 # 1. 在本地导出文件:
2 mysqldump -h9.145.42.250 -ujob_r -pjob_tmp_r --default-character-set=utf8 --
```

tail -n 15 查看导出文件:

```
9.134.238.73 (1) x
媒',10243101,'389885604','2020-03-20 07:38:09'),(1507017,'鱼子酱',117,'2211378696','2020-03-20 07:51:11'),(1507018,'星影传媒',219,'2330231981','2020-03-20 09:04:22'),
(1507019,'兵仙传媒',10242471,'76842216','2020-03-20 09:10:08'),(1507020,'星影传媒',219,'2327730201','2020-03-20 09:45:10'),(1507021,'柏橡文化',10243876,'1251858139','
2020-03-20 10:19:48'),(1507022,'鱼子酱',117,'522055883','2020-03-20 10:57:35'),(1507023,'壹智播客',10244116,'3258356133','2020-03-20 11:04:18'),(1507024,'星影传媒',21
9,'1584863763','2020-03-20 12:32:48'),(1507025,'星影传媒',219,'2332509614','2020-03-20 14:02:34'),(1507026,'柏橡文化',10243876,'1482055460','2020-03-21 00:23:43'),(15
07027,'柏橡文化',10243876,'781224162','2020-03-21 03:00:18'),(1507028,'柏橡文化',10243876,'3552148853','2020-03-21 04:02:43'),(1507029,'柏橡文化',10243876,'256868782'
,'2020-03-21 05:54:55'),(1507030,'柏橡文化',10243876,'3297905473','2020-03-21 06:31:36'),(1507031,'牧之传媒',10241617,'2801004553','2020-03-21 07:31:08'),(1507032,'星
影传媒',219,'2331128478','2020-03-22 05:23:43'),(1507033,'星影传媒',219,'1602823600','2020-03-22 06:37:35'),(1507034,'牧之传媒',10241617,'3308256541','2020-03-22 06:4
5:30'),(1507035,'优娱传媒',10243101,'3589109301','2020-03-22 07:51:57'),(1507036,'星影传媒',219,'2329530991','2020-03-22 14:22:22'),(1507037,'星影传媒',219,'232964965
0','2020-03-22 14:34:56'),(1507038,'星影传媒',219,'2329943167','2020-03-22 14:39:09'),(1507039,'柏橡文化',10243876,'2184069040','2020-03-23 02:32:43'),(1507040,'柏橡
文化',10243876,'2551036032','2020-03-23 02:33:14'),(1507041,'柏橡文化',10243876,'2549163935','2020-03-23 02:35:18'),(1507042,'星影传媒',219,'2241699365','2020-03-23 1
0:53:36'),(1507043,'天幕传媒',242,'519561808','2020-03-24 05:53:19'),(1507044,'牧之传媒',10241617,'1048767386','2020-03-24 09:33:49'),(1507045,'星影传媒',219,'1593845
928','2020-03-27 05:33:42'),(1507046,'星影传媒',219,'1570684201','2020-03-27 08:40:06'),(1507047,'星影传媒',219,'2327815789','2020-03-27 14:27:09'),(1507048,'柏橡文化'
,10243876,'2459250239','2020-03-28 07:19:17'),(1507049,'柏橡文化',10243876,'1730945836','2020-03-28 10:47:41'),(1507050,'柏橡文化',10243876,'3029154677','2020-03-30
10:04:23'),(1507051,'星影传媒',219,'2729700439','2020-03-31 14:20:55'),(1507052,'星影传媒',219,'1803637076','2020-04-02 09:11:46'),(1507053,'星影传媒',219,'2330792684
```

```
1 # 2. Linux命令计算文件MD5
2 md5sum union_operator_info.csv
```

结果:

```
[root@VM-238-73-centos ~]# md5sum union_operator_info.csv
e3ce74e3355a3d96398fe2c9d6cda709 union_operator_info.csv
```

```
1 -- 3. 去重sql
2
3 -- 方法一:
4 delete from union_operator_info where id not in (
5 select min(id) from union_operator_info group by union_name having count(unic
```

```

6         );
7
8  -- 方法二：聚合函数
9  select *, rank() over (partition by union_name order by id asc) as "去重" fro

```

select操作结果：

```

|      277 | 鸣扬传媒          | 10243164 | 307435037 | 2020-04-21 17:32:51 |
|      308 | 鸿运娱乐          | 10241557 | 57191708  | 2020-06-22 17:05:04 |
+-----+-----+-----+-----+-----+
34 rows in set (2.50 sec)

```