

北京邮电大学

数据库系统原理



第一次实验

姓 名 郑毓恒
学 院 计算机学院
专 业 计算机科学与技术
班 级 2020211302
学 号 2020211262
任课教师 邓芳

2022 年 10 月

实验一 GaussDB(for openGauss)数据库创建与维护实验

实验目的

1. 通过对 GAUSSDB(FOR OPENGauss)数据库创建与访问：
 - 1.1. 了解华为云分布式数据库 GAUSSDB(FOR OPENGauss)的软件环境和创建方法；
 - 1.2. 掌握并熟悉 GAUSSDB(FOR OPENGauss)数据库软件的使用方法；
 - 1.3. 掌握并熟悉 GAUSSDB(FOR OPENGauss)数据库软件的构成和相关工具；
 - 1.4. 通过 GAUSSDB(FOR OPENGauss)数据库软件的使用，深入理解数据库系统的基本概念。
2. 通过创建 GAUSSDB(FOR OPENGauss)数据库及进行相应的维护，了解并掌握 GaussDB(for openGauss)数据库的创建与维护的不同方法和途径，进而通过这一具体的数据库理解实际数据库所包含的各要素。

实验平台及环境

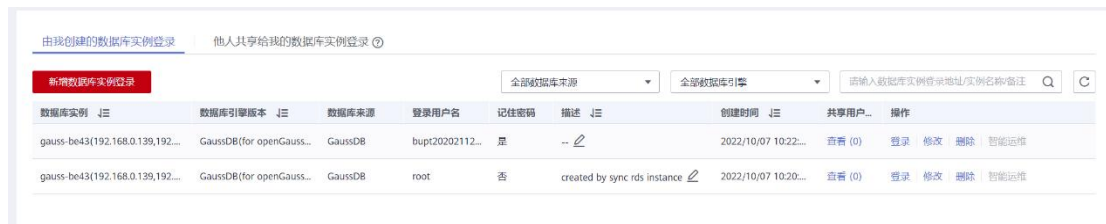
平台：华为云数据库数据管理服务 DAS 面向个体开发工具

环境：GaussDB(for openGauss)数据库，语言 PostgreSQL

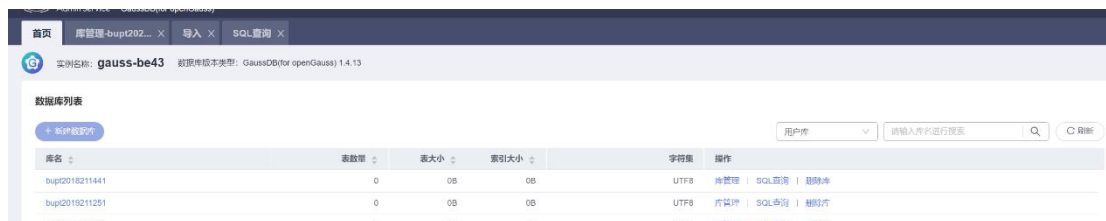
实验内容

1. 熟悉 GaussDB(for openGauss)数据库的创建过程；
2. 创建一个名为“疫情数据”的数据库；
3. 删除“疫情数据”数据库。

实验步骤



首先登录华为云，进入数据管理服务 DAS 个人开发工具，登录先前创建的数据库实例。

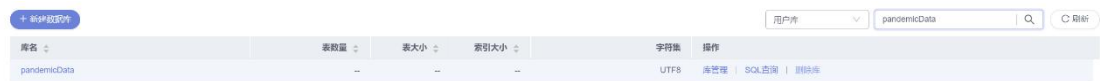


登录实例后，点击左上角新建数据库。



新建数据库，由于数据库名称只能为英文，因此使用'pandemicData'代替。字符集选择 UTF8, Template 选择 template0, DBCOMPATIBILITY 选择 PostgreSQL。点击确定完成创建。创建完毕后，再选择该数据库，进行删除。

实验结果及分析



The screenshot shows a database management interface with a search bar at the top right containing 'pandemicData'. Below the search bar is a table with columns: 库名 (Database Name), 表数量 (Table Count), 表大小 (Table Size), 索引大小 (Index Size), 字符集 (Character Set), and 操作 (Operations). The table contains one entry for 'pandemicData' with dashes in the first four columns and 'UTF8' in the fifth. Navigation links like '库管理' (Database Management), 'SQL查询' (SQL Query), and '删除库' (Delete Database) are visible at the bottom of the table.

库名	表数量	表大小	索引大小	字符集	操作
pandemicData	--	--	--	UTF8	库管理 SQL查询 删除库

在所有数据库中搜索 `pandemicData`，可以找到刚刚创建的数据库，证明数据创建成功。



点击删除库再点击‘是’，将该数据库删除。

实验小结

本次实验内容简单，只是简单的数据库创建和删除，并没有遇到什么问题。但是，从本次实验了解到华为云数据管理系统可以创建和使用的数据库类型。

实验二 数据库表的创建与维护实验

实验目的

1. 通过进行数据库表的建立操作，熟悉并掌握 GAUSSDB(FOR OPENGAUSS)数据库表的建立方法，理解关系型数据库表的结构，巩固 PostgreSQL 中关于数据库表的建立语句；
2. 通过进行数据库表数据的增加、删除和插入等维护操作，熟悉并掌握 GAUSSDB(FOR OPENGAUSS)数据库数据的操作方法，巩固 PostgreSQL 中关于数据维护的语句。

实验平台及环境

平台：华为云数据库数据管理服务 DAS 面向个体开发工具

环境：GaussDB(for openGauss)数据库，语言 PostgreSQL

实验内容

建立相应的表并熟悉基本操作，例如建表、对表进行增、删、改、查。

实验步骤

1. 熟悉课程实验背景知识；
2. 使用 GAUSSDB(FOR OPENGAUSS)数据库软件创建相应的表；
3. 将提供的数据导入各表，掌握 GAUSSDB(FOR OPENGAUSS)数据库数据导入的方法；
注意：
 - a. 表中空列的处理；
 - b. 表结构与数据类型的匹配。
4. 修改“病例基本信息”表数据，增加名为“备注”的列,数据类型为 varchar() 型；
5. 修改“病例基本信息”表数据，将“备注”列的数据类型改为 int；
6. 修改“病例基本信息”表数据，删除“备注”列；
7. 删除“病例基本信息”数据表。

实验结果及分析

```
CREATE TABLE 病例基本信息(  
  病例号    INT,  
  省        VARCHAR (20),  
  市        VARCHAR (20),  
  区        VARCHAR (20),  
  日期      VARCHAR (10),  
  性别      VARCHAR (10),  
  年龄      INT,  
  患者信息  VARCHAR (1000),  
  其它信息  VARCHAR (1000),  
  信息来源  VARCHAR (100)  
);
```

输入以上 SQL 语句，创建“病例基本信息”表。除了病例号和年龄设为 int 类型，其他设为 varchar，方便导入数据表格。



查看库中的表，可以看到“病例基本信息”表。点击查看，每列名称与 SQL 语句中设置的一致。

新建任务

导入类型

sqlCSV

文件来源

上传文件从OBS中选择

附件存放位置 ②

obs0630bupt

没有OBS桶? 创建OBS桶

选择附件

+

点击或将文件拖动到此处后上传文件 (.csv)

🕒 病例基本信息表_1665504352122.csv

最大不能超过1GB, 且只能上传一个附件

数据库

bupt2020211262

表

"bupt2020211262"."病例基本信息"

数据位置

第1行为属性第1行为数据

字符集

UTF8GBK自动检测

选项

☐ 忽略报错,即SQL执行失败时跳过

☒ 导入完成后删除上传的文件

☒ 清空表,执行导入前对相应的表执行Truncate操作

备注

创建导入任务取消

选择以上设置，将病例基本信息数据表格导入。

	创建时间	导入文件类型	库表	文件名	任务状态	执行时间	导入成功(行)	是否忽略报错	进度	备注	操作
bc	2022-10-12 00:28:24	CSV	目标库: bupt2020211262, 目标Schema: bupt2020211262, 目标表: 病例基本信息	病例基本信息表_166550562224.csv	🟢 已完成	1秒	1187	否	<div>100%</div>		查看详情
91	2022-10-12 00:26:16	CSV	目标库: bupt2020211262, 目标Schema: bupt2020211262, 目标表: 病例基本信息	病例基本信息表_1665505511231.csv	🔴 执行异常	2秒	296	否	<div>21.7%</div>		查看详情

导入成功。

当前所在库: bupt202011262 切换库 | 192.168.0.138:192.168.0.206:192.168.0.161:8000 | 字符集: UTF8 SQL 脚本 元数据管理

Schema列表 对象列表 元数据管理 打开表: 病例基本信息 X

Where条件 复制行 复制列 列设置

病例号	省	市	区	日期	性别	年龄	患者信息
1	河北省	石家庄市	藁城区	2021/1/20	女	54	确诊病例1: 女, 54岁, 藁城区
2	河北省	石家庄市	藁城区	2021/1/20	女	39	确诊病例2: 女, 39岁, 藁城区
3	河北省	石家庄市	藁城区	2021/1/20	女	84	确诊病例3: 女, 84岁, 藁城区
4	河北省	石家庄市	藁城区	2021/1/20	男	3	确诊病例4: 男, 3岁, 藁城区
5	河北省	石家庄市	藁城区	2021/1/20	女	5	确诊病例5: 女, 5岁, 藁城区
6	河北省	石家庄市	正定县	2021/1/20	女	33	确诊病例6: 女, 33岁, 正定县
7	河北省	石家庄市	藁城区	2021/1/20	女	49	确诊病例7: 女, 49岁, 藁城区
8	河北省	石家庄市	藁城区	2021/1/20	女	14	确诊病例8: 女, 14岁, 藁城区
9	河北省	石家庄市	长安区	2021/1/20	男	37	确诊病例9: 男, 37岁, 长安区
10	河北省	石家庄市	藁城区	2021/1/20	男	49	确诊病例10: 男, 49岁, 藁城区
11	河北省	石家庄市	藁城区	2021/1/20	女	67	确诊病例11: 女, 67岁, 藁城区
12	河北省	石家庄市	藁城区	2021/1/20	女	69	确诊病例12: 女, 69岁, 藁城区
13	河北省	石家庄市	长安区	2021/1/20	女	20	确诊病例13: 女, 20岁, 长安区

当前第 1 页 上一页 下一页 50 条/页 GO 刷新 单行详情

打开表格可见以上内容，数据正确导入。

```
ALTER TABLE 病例基本信息 ADD 备注 VARCHAR(1000);
```

使用以上语句，给“病例基本信息”表加入类型为 varchar(100)的备注列。

Where条件 复制行 复制列 列设置

年龄	患者信息	其它信息	信息来源	备注
54	确诊病例1: 女, 54岁, 藁城区	暂无	河北日报	
39	确诊病例2: 女, 39岁, 藁城区	暂无	河北日报	
84	确诊病例3: 女, 84岁, 藁城区	暂无	河北日报	
3	确诊病例4: 男, 3岁, 藁城区	暂无	河北日报	
5	确诊病例5: 女, 5岁, 藁城区	暂无	河北日报	
33	确诊病例6: 女, 33岁, 正定县	暂无	河北日报	
49	确诊病例7: 女, 49岁, 现住新	暂无	河北日报	
14	确诊病例8: 女, 14岁, 藁城区	暂无	河北日报	
37	确诊病例9: 男, 37岁, 现住长	暂无	河北日报	
49	确诊病例10: 男, 49岁, 藁城区	暂无	河北日报	
67	确诊病例11: 女, 67岁, 藁城区	暂无	河北日报	
69	确诊病例12: 女, 69岁, 藁城区	暂无	河北日报	
20	确诊病例13: 女, 20岁, 现住长	暂无	河北日报	

刷新 单行详情

打开病例基本信息表，可以看到新增的“备注”列，新增成功。

```
ALTER TABLE 病例基本信息
```

```
MODIFY 备注 INT;
```

执行以上 SQL 语句，将“备注”列更改为 int 数据类型。

ALTER TABLE 病例基本信息

DROP COLUMN 备注;

执行以上 SQL 语句，将“备注”列删除。

Where条件 复制行 复制列 列设置					
	性别	年龄	患者信息	其它信息	信息来源
	女	54	确诊病例1: 女, 54岁, 藁城区	暂无	河北日报
	女	39	确诊病例2: 女, 39岁, 藁城区	暂无	河北日报
	女	84	确诊病例3: 女, 84岁, 藁城区	暂无	河北日报
	男	3	确诊病例4: 男, 3岁, 藁城区	暂无	河北日报
	女	5	确诊病例5: 女, 5岁, 藁城区	暂无	河北日报
	女	33	确诊病例6: 女, 33岁, 正定县	暂无	河北日报
	女	49	确诊病例7: 女, 49岁, 现住新	暂无	河北日报
	女	14	确诊病例8: 女, 14岁, 藁城区	暂无	河北日报
	男	37	确诊病例9: 男, 37岁, 现住长	暂无	河北日报
	男	49	确诊病例10: 男, 49岁, 藁城区	暂无	河北日报
	女	67	确诊病例11: 女, 67岁, 藁城区	暂无	河北日报
	女	69	确诊病例12: 女, 69岁, 藁城区	暂无	河北日报
	女	20	确诊病例13: 女, 20岁, 现住长	暂无	河北日报

打开病例基本信息表，可见“备注”列已被删除。

DROP TABLE 病例基本信息;

执行以上 SQL 语句，将“病例基本信息”表删除。

Schema列表 对象列表 元数据列表					
对象列表数据库DAS验证时采集信息，最近一次采集时间: 2022-10-11 17:28:37 立即采集 历史记录					
表	Schema: bup2020211202	按条件进行过滤			
视图	表名	行数	表大小	索引大小	字符数 操作
存储过程	暂无数据				
触发器					
序列					

查看数据库中的表，没有“病例基本信息”，删除成功。

实验小结

实验在导入过程中遇到了问题，首先是导入的表格中的列和数据库中的目标表格的列要一致，不仅是它们的名称，还有它们的数据类型。比如，一开始在“患者信息”列只设置了 `varchar(100)`，长度只有 100 个字符，大约为 50 个中文字符。然而，在原数据表格中，患者信息存在长度超过 50 个中文字符的项，因此导入失败。将该列的数据类型改为 `varchar(1000)`，导入成功。从这点学习到，假如不能确定原数据表格的数据格式、类型或者规范性，可以先将数据类型设为足够长的 `varchar` 或表示范围够大的其他数据类型。

导入表格完成后，剩余的删改列和删除表操作没有遇到问题。从本次实验学习到利用 SQL 语句对表格进行基本的创建、修改和删除操作，以及如何将外部数据导入进数据库中，为后续实验建立了基础。

实验三 数据查询实验

实验目的

通过对实验二建立的数据库关系表的各种查询的操作，加深对 SQL 语言和 PostgreSQL 查询语言的了解，掌握相关查询语句的语法及使用方法。

实验平台及环境

平台：华为云数据库数据管理服务 DAS 面向个体开发工具

环境：GaussDB(for openGauss)数据库，语言 PostgreSQL

实验内容

1.单表查询

- 1-1 查询国内确诊病例基本信息的所有信息来源。
- 1-2 给出河南省、西藏自治区、台湾省的英文名称和人口数。
- 1-3 查询 2021 年 1 月 20 日各省现有确诊病例数据，按现有确诊病例数降序排列输出。
- 1-4 顺义区中风险地区的数量。
- 1-5 计算截至 2021 年 1 月 20 日全国累计确诊病例数。
- 1-6 查询 1005 号病例确诊后，其所在市新增的所有确诊病例。
- 1-7 在“病例基本信息表”中查询石家庄市在 2021 年 1 月 11 日当天以及之前的所有 60 岁以上的患者信息。
- 1-8 统计截止到 2020 年 12 月 30 日美国累计确诊病例数最多的 10 个州。
- 1-9 统计截至 2021 年 1 月 20 号中国发病率最高的人群(人群按照年龄划分，儿童<18，18<=青壮年<60，老年>=60)。

2.多表查询：

- 2-1 借助病例行程信息粗略查询曾去过“源升品质生活坊”的所有患者的基本信息。
- 2-2 根据病例行程信息表和病例基本信息表，查询行程信息中存在“家庭聚餐”的病例被确诊的日期。
- 2-3 对比中美两国累计确诊病例数，输出格式为(日期，中国累计确诊，美国累计确诊)。
- 2-4 计算截止到 2021 年 1 月 20 日，美国有些县的累计确诊是同一个州的其他县的 2 倍或以上，列出这些县，以及他们所在的州和他们的累计确诊。
- 2-5 计算世界上人口数排名前 10 位的国家地区。
- 2-6 列出美国人口超千万的大州中，截至 2021 年 1 月 20 日新冠肺炎疫情死亡率超过 2%的州。
- 2-7 截至 2021 年 1 月 20 日，河北省哪些区出现了新冠确诊病例但不属于中高风险地区。
- 2-8 在病例行程信息表的基础上根据病例基本信息表，查询河北省病例的全部信息。

3. 嵌套查询：

3-1 查询披露的确诊患者信息中年龄最大的患者，输出其基本信息。(未注明年龄的患者不进行比较)。

3-2 查询 2020 年 12 月份新增确诊患者最多的城市。

3-3 结合“全国各省参考信息表”和“病例基本信息表”给出没有新增确诊病例或未披露病例信息的省份。

3-4 2021 年 1 月 20 日全国中高风险地区所在省中，哪些省在 1 月 20 日没有新增确诊信息披露。

3-5 根据病例基本信息表查询一月份国内新增患者病例最多的城市。

3-6 查询除中美两国以外的其余国家中，进入 2021 年以来单日新增确诊病例始终不低于一万例的国家。

实验步骤

```
CREATE TABLE 病例基本信息(  
    病例号      INT,  
    省          VARCHAR(50),  
    市          VARCHAR(50),  
    区          VARCHAR(50),  
    日期        VARCHAR(10),  
    性别        VARCHAR(10),  
    年龄        INT,  
    患者信息    VARCHAR(1000),  
    其它信息    VARCHAR(1000),  
    信息来源    VARCHAR(100)  
);
```

```
CREATE TABLE 病例行程信息(  
    行程号      INT,  
    病例号      INT,  
    日期信息    VARCHAR(1000),  
    行程信息    VARCHAR(5000)  
);
```

```
CREATE TABLE 参考信息(  
    组合码      VARCHAR(100),  
    国家        VARCHAR(100),  
    省州        VARCHAR(500),  
    市县        VARCHAR(500),  
    纬度        FLOAT,  
    经度        FLOAT,  
    人口数      INT  
);
```

```
CREATE TABLE 各国疫情数据统计(  
    国家        VARCHAR(100),  
    日期        VARCHAR(10),  
    新增确诊    INT,  
    累计确诊    INT,  
    死亡        INT,  
    治愈        INT,  
    境外输入    INT,  
    境外输入日期 VARCHAR(10),  
    境外输入数量 INT,  
    境外输入来源 VARCHAR(100),  
    境外输入日期 VARCHAR(10),  
    境外输入数量 INT,  
    境外输入来源 VARCHAR(100)
```

```

    日期          VARCHAR(100),
    国家          VARCHAR(100),
    省州          VARCHAR(500),
    累计确诊      INT,
    累计治愈      INT,
    累计死亡      INT
);

CREATE TABLE 美国各州县确诊与死亡数统计(
    日期          VARCHAR(100),
    州            VARCHAR(100),
    县            VARCHAR(100),
    累计确诊      INT,
    累计死亡      INT
);

CREATE TABLE 全国城市风险等级(
    省            VARCHAR(100),
    市            VARCHAR(100),
    区            VARCHAR(100),
    地址详情      VARCHAR(1000),
    风险等级      VARCHAR(20)
);

CREATE TABLE 全国各省参考信息(
    中文名称      VARCHAR(100),
    英文名称      VARCHAR(100),
    组合码        VARCHAR(100),
    人口数        INT,
    省会城市      VARCHAR(100),
    纬度          FLOAT,
    经度          FLOAT
);

CREATE TABLE 全国各省累计数据统计(
    日期          VARCHAR(100),
    省            VARCHAR(100),
    累计确诊      INT,
    累计治愈      INT,
    累计死亡      INT
);

```

首先执行以上 SQL 语句，为要导入的所有数据表格建立表，然后将原始表格导入。

创建时间	导入文件类型	库表	文件名	任务状态	执行时间	导入成功(行)	是否新建表	进度	备注	操作
2022-10-12 01:30:38	CSV	目标库: bupd2020211262, 目标Schema: bupd2020211262, 目标表: 全国各省累计数据统计	全国各省累计数据统计_1665509421993.csv	已完成	1秒以内	2040	否	100%		查看详情
2022-10-12 01:30:11	CSV	目标库: bupd2020211262, 目标Schema: bupd2020211262, 目标表: 全国各省参考信息	全国各省参考信息_1665509336355.csv	已完成	1秒以内	34	否	100%		查看详情
2022-10-12 01:29:42	CSV	目标库: bupd2020211262, 目标Schema: bupd2020211262, 目标表: 全国城市风险等级	全国城市风险等级_1665509367361.csv	已完成	1秒以内	68	否	100%		查看详情
2022-10-12 01:29:05	CSV	目标库: bupd2020211262, 目标Schema: bupd2020211262, 目标表: 美国2019年确诊治愈死亡人数统计	美国2019年确诊治愈死亡人数统计_1665509330487.csv	已完成	8秒	150802	否	100%		查看详情
2022-10-12 01:28:37	CSV	目标库: bupd2020211262, 目标Schema: bupd2020211262, 目标表: 全国疫情数据统计	全国疫情数据统计_1665509311819.csv	已完成	1秒	14409	否	100%		查看详情

此处显示全部导入成功，可以开始实验。

对实验内容的每一步，分别执行以下 SQL 语句完成任务。

1. 单表查询

1-1 查询国内确诊病例基本信息的所有信息来源。

```
SELECT 信息来源
FROM 病例基本信息;
```

1-2 给出河南省、西藏自治区、台湾省的英文名称和人口数。

```
SELECT 中文名称, 英文名称, 人口数
FROM 全国各省参考信息
WHERE 中文名称 = '河南省' or 中文名称 = '西藏自治区' or 中文名称 = '台湾省';
```

1-3 查询 2021 年 1 月 20 日各省现有确诊病例数据，按现有确诊病例数降序排列输出。

```
SELECT 省, CAST(累计确诊-累计死亡-累计治愈 AS INT) AS 现有确诊
FROM 全国各省累计数据统计
WHERE 日期 = '2021/1/20'
ORDER BY 现有确诊 DESC
```

1-4 顺义区中风险地区的数量。

```
SELECT count(*)
FROM 全国城市风险等级
WHERE 区 = '顺义区' and 风险等级 = '中风险地区';
```

1-5 计算截至 2021 年 1 月 20 日全国累计确诊病例数。

```
SELECT SUM(累计确诊)
FROM 全国各省累计数据统计
WHERE 日期 = '2021/1/20';
```

1-6 查询 1005 号病例确诊后，其所在市新增的所有确诊病例。
<pre> ALTER TABLE 病例基本信息 MODIFY 日期 DATE; SELECT * FROM 病例基本信息 AS A, (SELECT 省, 市 FROM 病例基本信息 WHERE 病例号 = 1005)AS B WHERE A.日期 > '2021-01-06' and A.省 = B.省 and A.市 = B.市; </pre> <p>*创建“病例基本信息”时，将日期的类型设为 varchar()，因此要先改为 DATE 类型才能进行比较。</p>
1-7 在“病例基本信息表”中查询石家庄市在 2021 年 1 月 11 日当天以及之前的所有 60 岁以上的患者信息。
<pre> SELECT * FROM 病例基本信息 WHERE 日期 <= '2021-01-11' and 省 = '河北省' and 市 = '石家庄市' and 年龄 > 60; </pre>
1-8 统计截止到 2020 年 12 月 30 日美国累计确诊病例数最多的 10 个州。
<pre> ALTER TABLE 美国各州县确诊与死亡数统计 MODIFY 日期 date; SELECT 州, SUM(累计确诊) FROM 美国各州县确诊与死亡数统计 WHERE 日期 = '2020-12-30' GROUP BY 州 ORDER BY SUM(累计确诊) DESC LIMIT 10; </pre> <p>*创建“美国各州县确诊与死亡数统计”时，将日期的类型设为 varchar()，因此要先改为 DATE 类型才能进行比较。</p>
1-9 统计截至 2021 年 1 月 20 号中国发病率最高的人群（人群按照年龄划分，儿童<18，18<=青壮年<60，老年>=60）。
<pre> ALTER TABLE 病例基本信息 ADD COLUMN 人群 VARCHAR(10); UPDATE 病例基本信息 SET 人群 = '儿童' WHERE 年龄 < 18; UPDATE 病例基本信息 SET 人群 = '青壮年' WHERE 年龄 >= 18 and 年龄 < 60; </pre>

```

UPDATE 病例基本信息
SET 人群 = '老年'
WHERE 年龄 >= 60;

SELECT 人群, COUNT(人群)
FROM 病例基本信息
WHERE 日期 <= '2021-01-20' AND 人群 IS NOT NULL
GROUP BY 人群
ORDER BY COUNT(人群) DESC
LIMIT 1;

```

*先给“病例基本信息”表加入“人群”列，并更新相应数据。

2.多表查询

2-1 借助病例行程信息粗略查询曾去过“源升品质生活坊”的所有患者的基本信息。

```

SELECT *
FROM 病例基本信息
WHERE 病例号 IN (
    SELECT DISTINCT 病例号
    FROM 病例行程信息
    WHERE 行程信息 LIKE '%源升品质生活坊%');

```

2-2 根据病例行程信息表和病例基本信息表，查询行程信息中存在“家庭聚餐”的病例被确诊的日期。

```

SELECT 日期
FROM 病例基本信息
WHERE 病例号 IN (
    SELECT DISTINCT 病例号
    FROM 病例行程信息
    WHERE 行程信息 LIKE '%家庭聚餐%');

```

2-3 对比中美两国累计确诊病例数，输出格式为(日期，中国累计确诊，美国累计确诊)。

```

SELECT C.日期, C.中国累计确诊, A.美国累计确诊
FROM
    (
        SELECT 日期, SUM(累计确诊) AS 中国累计确诊
        FROM 全国各省累计数据统计
        GROUP BY 日期
        ORDER BY 日期
    ) AS C,
    (
        SELECT 日期, SUM(累计确诊) AS 美国累计确诊

```



```

FROM 美国各州县确诊与死亡数统计
GROUP BY 日期
ORDER BY 日期
) AS A
WHERE C.日期 = A.日期
ORDER BY 日期

```

2-4 计算截止到 2021 年 1 月 20 日，美国有些县的累计确诊是同一个州的其他县的 2 倍或以上，列出这些县，以及他们所在的州和他们的累计确诊。

```

SELECT B.州, B.县, B.县累计确诊
FROM
(
  SELECT 州, SUM(累计确诊) AS 州累计确诊
  FROM 美国各州县确诊与死亡数统计
  WHERE 日期 = '2021-01-20'
  GROUP BY 州
) AS A,
(
  SELECT 州, 县, 累计确诊 AS 县累计确诊
  FROM 美国各州县确诊与死亡数统计
  WHERE 日期 = '2021-01-20'
) AS B
WHERE A.州 = B.州 AND B.县累计确诊 >= 2 * (A.州累计确诊 - B.县累计确诊)

```

2-5 计算世界上人口数排名前 10 位的国家地区。

```

SELECT 国家, SUM(人口数) AS 人口总数
FROM 参考信息
WHERE 人口数 IS NOT NULL AND 省州 IS NULL
GROUP BY 国家
ORDER BY 人口总数 DESC
LIMIT 10;

```

2-6 列出美国人口超千万的大州中，截至 2021 年 1 月 20 日新冠肺炎疫情死亡率超过 2% 的州。

```

SELECT A.州, B.人口数, CAST(A.累计死亡/B.人口数 AS DECIMAL(10, 2))
AS 死亡率
FROM
(
  SELECT 州, SUM(累计死亡) AS 累计死亡
  FROM 美国各州县确诊与死亡数统计
  WHERE 日期 = '2021-01-20'
  GROUP BY 州
) AS A,

```

```
(
    SELECT 省州, 人口数
    FROM 参考信息
    WHERE 国家 = 'US' AND 省州 IS NOT NULL AND 市县 IS NULL
)AS B
WHERE B.人口数 >= 10000000 AND A.州 = B.省州 AND A.累计死亡/B.人口
数 > 0.02;
```

2-7 截至 2021 年 1 月 20 日，河北省哪些区出现了新冠确诊病例但不属于中高风险地区。

```
SELECT A.市, A.区
FROM
(
    SELECT 市, 区
    FROM 病例基本信息
    WHERE 省 = '河北省'
)AS A
WHERE NOT EXISTS
(
    SELECT 市, 区
    FROM 全国城市风险等级 AS B
    WHERE 省 = '河北省' AND A.市 = B.市 AND A.区 = B.区
);
```

2-8 在病例行程信息表的基础上根据病例基本信息表，查询河北省病例的全部信息。

```
SELECT A.*, B.*
FROM 病例基本信息 AS A, 病例行程信息 AS B
WHERE A.病例号 = B.病例号 AND A.省 = '河北省';
```

3. 嵌套查询

3-1 查询披露的确诊患者信息中年龄最大的患者，输出其基本信息。(未注明年龄的患者不进行比较)。

```
SELECT A.*
FROM 病例基本信息 AS A
WHERE A.年龄 >= ALL (
    SELECT MAX(年龄)
    FROM 病例基本信息
)
```

3-2 查询 2020 年 12 月份新增确诊患者最多的城市。

```
SELECT A.*
FROM
(
```

```

SELECT 市, COUNT(*) AS 十二月新增确诊数
FROM 病例基本信息
WHERE 日期 >= '2020-12-01' AND 日期 < '2021-01-01'
GROUP BY 市
)AS A
ORDER BY A.十二月新增确诊数 DESC
LIMIT 1;

```

3-3 结合“全国各省参考信息表”和“病例基本信息表”给出没有新增确诊病例或未披露病例信息的省份。

```

SELECT *
FROM
(
    SELECT DISTINCT 中文名称 AS 省
    FROM 全国各省参考信息
)AS A
WHERE NOT EXISTS(
    SELECT DISTINCT 省
    FROM 病例基本信息 AS B
    WHERE A.省 = B.省
)

```

3-4 2021 年 1 月 20 日全国中高风险地区所在省中，哪些省在 1 月 20 日没有新增确诊信息披露。

```

SELECT *
FROM
(
    SELECT DISTINCT 省
    FROM 全国城市风险等级
)AS A
WHERE NOT EXISTS(
    SELECT DISTINCT 省
    FROM 病例基本信息 AS B
    WHERE A.省 = B.省 AND 日期 = '2021-01-20'
)

```

3-5 根据病例基本信息表查询一月份国内新增患者病例最多的城市。

```

SELECT A.*
FROM
(
    SELECT 市, COUNT(*) AS 一月新增确诊数
    FROM 病例基本信息
    WHERE 日期 >= '2021-01-01'
    GROUP BY 市
)

```

```
)AS A
ORDER BY A.一月新增确诊数 DESC
LIMIT 1
```

3-6 查询除中美两国以外的其余国家中, 进入 2021 年以来单日新增确诊病例始终不低于一万例的国家。

```
ALTER TABLE 各国疫情数据统计
MODIFY 日期 DATE

SELECT A.国家
FROM
(
    SELECT DISTINCT 国家
    FROM 各国疫情数据统计
    WHERE 国家 <> 'US' AND 国家 <> 'China'
)AS A
WHERE 国家 NOT IN
(
    SELECT D.国家
    FROM (
        SELECT C.日期, C.国家, CAST(C.累计确诊 - B.累计确诊 AS INT) AS 单日
        新增确诊
        FROM (
            SELECT 日期, 国家, SUM(累计确诊) AS 累计确诊
            FROM 各国疫情数据统计
            GROUP BY 日期, 国家
        )AS B,
        (
            SELECT 日期, 国家, SUM(累计确诊) AS 累计确诊
            FROM 各国疫情数据统计
            GROUP BY 日期, 国家
        )AS C
        WHERE B.日期 >= '2020-12-31' AND C.日期 - B.日期 = 1 AND C.国家 =
        B.国家 )AS D
        WHERE D.单日新增确诊 < 10000
    )
```

*创建“各国疫情数据统计”时, 将日期的类型设为 varchar(), 因此要先改为 DATE 类型才能进行比较。

实验结果及分析

以上 SQL 语句的执行结果如下。

1-1 查询国内确诊病例基本信息的所有信息来源。

以下是SELECT 信息来源 FROM 病例基本信息表的执行结果集

复制行 复制列 列设置

	信息来源
1	河北日报
2	河北日报
3	河北日报
4	河北日报
5	河北日报
6	河北日报
7	...

最多显示 50 行

刷新 单行详情

1-2 给出河南省、西藏自治区、台湾省的英文名称和人口数。

以下是SELECT 中文名称, 英文名称, 人口数 FROM 全国各省参考信息 WHERE 中文名称 = '河南省' or ...

复制行 复制列 列设置

	中文名称	英文名称	人口数
1	河南省	Henan	96050000
2	台湾省	Taiwan	23816775
3	西藏自治区	Tibet	3440000

最多显示 50 行

刷新 单行详情

1-3 查询 2021 年 1 月 20 日各省现有确诊病例数据，按现有确诊病例数降序排列输出。

以下是SELECT 省, CAST(累计确诊-累计死亡-累计治愈 AS INT) AS 现有确诊 FROM 全国各省累计数...

复制行 复制列 列设置

	省	现有确诊
1	河北省	829
2	香港特别行政区	766
3	黑龙江省	259
4	吉林省	162
5	上海市	95
6	北京市	44
7	广东省	39
8	陕西省	26

1-4 顺义区中风险地区的数量。

SQL执行记录 消息 结果集1 X

以下是SELECT count(*) FROM 全国城市风险等级 WHERE 区 = '顺义区' and 风险等级 = '中风险地区'...

复制行 复制列 列设置

	count
1	4

最多显示 50 行

刷新 单行详情

1-5 计算截至 2021 年 1 月 20 日全国累计确诊病例数。

以下是SELECT SUM(累计确诊) FROM 全国各省累计数据统计 WHERE 日期 = '2021/1/20'的统计结果... 该表不可编辑。

	sum
1	98545

最多显示 50 行

刷新 单行详情

1-6 查询 1005 号病例确诊后，其所在市新增的所有确诊病例。

以下是SELECT * FROM 病例基本信息 WHERE 日期 > '2021-01-06' and 省 = '黑龙江省' and 市 = '黑... 该表不可编辑。

病例号	省	市	区	日期	性别	年龄
1	黑龙江省	黑河市	爱辉区	2021-01-07	女	5

最多显示 50 行

刷新 单行详情

1-7 在“病例基本信息表”中查询石家庄市在 2021 年 1 月 11 日当天以及之前的所有 60 岁以上的患者信息。

以下是SELECT * FROM 病例基本信息 WHERE 日期 <= '2021-01-11' and 省 = '河北省' and 市 = '石... 该表不可编辑。

病例号	省	市	区	日期	性别	年龄
1	河北省	石家庄市	藁城区	2021-01-11	女	7
2	河北省	石家庄市	藁城区	2021-01-11	女	7
3	河北省	石家庄市	藁城区	2021-01-11	女	6
4	河北省	石家庄市	藁城区	2021-01-11	女	7
5	河北省	石家庄市	藁城区	2021-01-11	男	7

最多显示 50 行

刷新 单行详情

1-8 统计截止到 2020 年 12 月 30 日美国累计确诊病例数最多的 10 个州。

以下是SELECT 州, SUM(累计确诊) FROM 美国各州确诊与死亡统计 WHERE 日期 = '2020-12-3... 该表不可编辑。

	州	sum
1	California	2291414
2	Texas	1754075
3	Florida	1306123
4	New York	963390
5	Illinois	955380
6	Ohio	690748
7	Georgia	654743
8	Pennsylvania	636061
9	Tennessee	580809
10	North Carolina	532830

最多显示 50 行

刷新 单行详情

1-9 统计截至 2021 年 1 月 20 号中国发病率最高的人群（人群按照年龄划分，儿童<18，18<=青壮年<60，老年>=60）。

以下是SELECT 人群, COUNT(人群) FROM 病例基本信息 WHERE 日期 <= '2021-01-20' AND 人群 != '...'. 该表不可编辑。

	人群	count
1	青壮年	696

最多显示 50 行

2-1 借助病例行程信息粗略查询曾去过“源升品质生活坊”的所有患者的基本信息。

以下是SELECT * FROM 病例基本信息 WHERE 病例号 IN (SELECT DISTINCT 病例号 FROM 病例...). 该表不可编辑。

	病例号	省	市	区	日期	性别	年龄
1	157	吉林省	通化市	东昌区	2021-01-18	女	7
2	167	吉林省	通化市	东昌区	2021-01-18	女	7
3	171	吉林省	通化市	东昌区	2021-01-18	男	7
4	568	吉林省	通化市	东昌区	2021-01-13	男	4

最多显示 50 行

2-2 根据病例行程信息表和病例基本信息表，查询行程信息中存在“家庭聚餐”的病例被确诊的日期。

以下是SELECT 日期 FROM 病例基本信息 WHERE 病例号 IN (SELECT DISTINCT 病例号 FROM 病例...). 该表不可编辑。

	日期
1	2021-01-13
2	2021-01-05
3	2021-01-04
4	2021-01-13

最多显示 50 行

2-3 对比中美两国累计确诊病例数，输出格式为(日期，中国累计确诊，美国累计确诊)。

以下是SELECT C.日期, C.中国累计确诊, A.美国累计确诊 FROM (SELECT 日期, SUM(累计确诊) AS...). 该表不可编辑。

	日期	中国累计确诊	美国累计确诊
1	2020/11/22	92117	12347905
2	2020/11/23	92212	12521898
3	2020/11/24	92297	12697801
4	2020/11/25	92403	12879677
5	2020/11/26	92489	12991818

最多显示 50 行

2-4 计算截止到 2021 年 1 月 20 日，美国有些县的累计确诊是同一个州的其他县

的 2 倍或以上，列出这些县，以及他们所在的州和他们的累计确诊。

以下是SELECT B.州, B.县, B.县累计确诊 FROM (SELECT 州, SUM(累计确诊) AS 州累计确诊 FROM ...) 该表不可编辑。

	州	县	县累计确诊
1	Grand Princess		103
2	Diamond Princess		49
3	Virgin Islands		2305
4	District of Columbia	District of Columbia	34403
5	Guam		7521
6	American Samoa		0
7	Hawaii	Honolulu	20231
8	Northern Mariana Islands		129

最多显示 50 行 刷新 单行详情

2-5 计算世界上人口数排名前 10 位的国家地区。

以下是SELECT 国家, SUM(人口数) AS 人口总数 FROM 参考信息 WHERE 人口数 IS NOT NULL AND ... 该表不可编辑。

	国家	人口总数
1	India	1380004385
2	US	329466283
3	Indonesia	273523621
4	Pakistan	220892331
5	Brazil	212559409
6	Nigeria	206129587
7	Bangladesh	164689383
8	Russia	145934460
9	Mexico	127792286

最多显示 50 行 刷新 单行详情

2-6 列出美国人口超千万的大州中，截至 2021 年 1 月 20 日新冠肺炎疫情死亡率超过 2%的州。

以下是SELECT A.州, B.人口数, CAST(A.累计死亡/B.人口数 AS DECIMAL(10, 2)) AS 死亡率 FROM (...) 该表不可编辑。

	州	人口数	死亡率
暂无数据			

最多显示 50 行 刷新 单行详情

2-7 截至 2021 年 1 月 20 日，河北省哪些区出现了新冠确诊病例但不属于中高风险地区。

以下是SELECT A.市, A.区 FROM (SELECT 市, 区 FROM 病例基本信息 WHERE 省 = '河北省') AS A ... 该表不可编辑。

	市	区
1	石家庄市	井陉县
2	石家庄市	裕华区
3	石家庄市	藁城
4	石家庄市	正定
5	南宫市	

最多显示 50 行 刷新 单行详情

2-8 在病例行程信息表的基础上根据病例基本信息表，查询河北省病例的全部信息。

以下是SELECT A.*, B.* FROM 病例基本信息 AS A, 病例行程信息 AS B WHERE A.病例号 = B.病例号... 该表不可编辑。

	病例号	省	市	区	日期	性别	年龄
1	1	河北省	石家庄市	藁城区	2021-01-20	女	5
2	1	河北省	石家庄市	藁城区	2021-01-20	女	5
3	1	河北省	石家庄市	藁城区	2021-01-20	女	5
4	1	河北省	石家庄市	藁城区	2021-01-20	女	5
5	10	河北省	石家庄市	藁城区	2021-01-20	男	4
6	10	河北省	石家庄市	藁城区	2021-01-20	男	4
7	10	河北省	石家庄市	藁城区	2021-01-20	男	4
8	100	河北省	石家庄市	藁城区	2021-01-18	男	8
9	100	河北省	石家庄市	藁城区	2021-01-18	男	8

最多显示 50 行 刷新 单行详情

3-1 查询披露的确诊患者信息中年龄最大的患者，输出其基本信息。(未注明年龄的患者不进行比较)。

以下是SELECT A.* FROM 病例基本信息 AS A WHERE A.年龄 >= ALL (SELECT MAX(年龄) FROM ...) 该表不可编辑。

病例号	省	市	区	日期	性别	年龄
1103	辽宁省	沈阳市	皇姑区	2020-12-29	女	95

最多显示 50 行 刷新 单行详情

3-2 查询 2020 年 12 月份新增确诊患者最多的城市。

以下是SELECT A.* FROM (SELECT 市, COUNT(*) AS 十二月新增确诊数 FROM 病例基本信息 WHERE ...) 该表不可编辑。

	市	十二月新增确诊数
1	大连市	38

最多显示 50 行 刷新 单行详情

3-3 结合“全国各省参考信息表”和“病例基本信息表”给出没有新增确诊病例或未披露病例信息的省份。

以下是SELECT * FROM (SELECT DISTINCT 中文名称 AS 省 FROM 全国各省参考信息) AS A WHERE ... 该表不可编辑。

	省
1	湖北省
2	河南省
3	台湾省
4	广西省
5	海南省
6	重庆市
7	福建省
8	安徽省

最多显示 50 行 刷新 单行详情

3-4 2021 年 1 月 20 日全国中高风险地区所在省中，哪些省在 1 月 20 日没有新增确诊信息披露。

以下是SELECT * FROM (SELECT DISTINCT 省 FROM 全国城市风险等级)AS A WHERE NOT EXI... ① 该表不可编辑。

	省
1	吉林省
2	辽宁省

最多显示 50 行

刷新 单行详情

3-5 根据病例基本信息表查询一月份国内新增患者病例最多的城市。

以下是SELECT A.* FROM (SELECT 市, COUNT(*)AS 一月新增确诊数 FROM 病例基本信息 WHERE... ① 该表不可编辑。

	市	一月新增确诊数
1	石家庄市	768

最多显示 50 行

刷新 单行详情

3-6 查询除中美两国以外的其余国家中，进入 2021 年以来单日新增确诊病例始终不低于一万例的国家。

以下是SELECT A.国家 FROM (SELECT DISTINCT 国家 FROM 每日新增确诊统计 WHERE 国家 <> 'US' AND 国家 < 2021/1/1... ① 该表不可编辑。

	国家
1	Russia
2	United Kingdom
3	Brazil

实验小结

本次实验主要以编写 SQL 代码完成特定查询任务为主，遇到的问题很多是有关这一方面的。例如，更改表的某一列的名称，在打完 ALTER TABLE 表名称之后，无论是输入 ALTER COLUMN 还是 MODIFY COLUMN 都执行失败，后来才知道只需要打 MODIFY 就能成功。系统并没有自动补全，当时以为并没有 MODIFY 这个关键字。又例如 SQL 要求选择的列中的，除了使用函数处理的以外的所有列，都需要使用 Group By，而有的情况并不需要。这里主要还是代码思路问题，错误使用了 SUM()、COUNT()等函数。

在本次实验，通过对数据库的实践操作和 SQL 的编写，对数据库原理和 SQL 语句的使用有了更深刻的认识，并掌握了更多使用技巧，为后续实验建立了基础。同时，也了解到 SQL 处理数据表的便利性，激发了我对数据库的兴趣，希望能在后续课程和实验中，学习到更多知识，并在往后利用于程序设计编写中。