

数据库系统概论实验报告

实验四：两种方式向数据表添加、导入和导出数据

陈羿羽 *

2021 年 5 月 14 日

摘要

本次实验的目标为：

1. 使用数据库管理系统 SQL Server 或者 MySQL 操作数据库。
2. 掌握创建、查看、修改和删除数据表的基本操作。

*陈羿羽：西南大学 2019 级计算机科学与技术 4 班 - 222019603193014

1 实验准备

1. 一台 PC 机或笔记本电脑
2. 能够上网

2 实验要求

1. 查资料观看数据库管理系 SQL Server 或者 MySQL 的数据表数据添加、导入和导出视频。
2. 使用视图方式和命令方式向数据表添加、导入和导出数据。
3. 学习了解数据表字段的基本类型。

3 实验内容

3.1 视图方式建立数据表

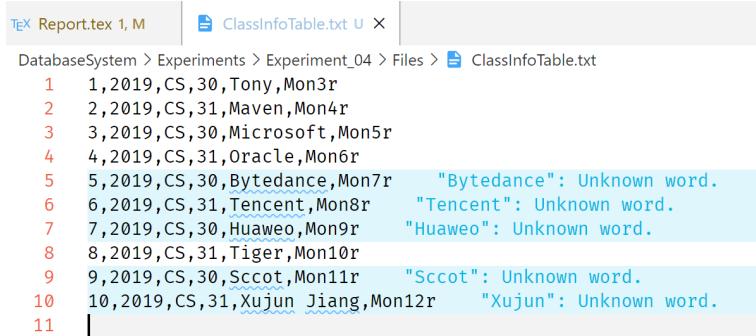
向实验三创建的数据表添加数据:

3.1.1 文件方式导入数据

自己各构造 10 条存在名字为“学生成绩.txt”和“班级信息.txt”的文本文件中，然后导入到数据库响应¹表中。(其中涉及姓名的信息要有一条和自己的姓名相关的信息)

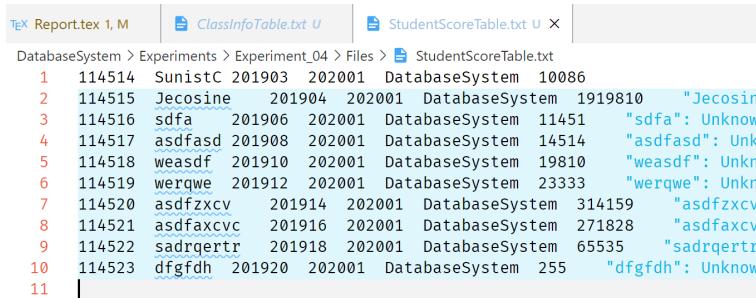
¹"响应表"是"相应"的表吗，并没有数据库响应表达玩意

此处我们构造了十条学生成绩信息记录与班级信息记录，如下图 1 和图 2 所示，其中学生成绩信息记录为制表符格式的文本文件，班级信息记录为逗号分隔符格式的文本文件。



```
TeX Report.tex 1, M | ClassInfoTable.txt U X
DatabaseSystem > Experiments > Experiment_04 > Files > ClassInfoTable.txt
1 1,2019,CS,30,Tony,Mon3r
2 2,2019,CS,31,Maven,Mon4r
3 3,2019,CS,30,Microsoft,Mon5r
4 4,2019,CS,31,Oracle,Mon6r
5 5,2019,CS,30,Bytedance,Mon7r "Bytedance": Unknown word.
6 6,2019,CS,31,Tencent,Mon8r "Tencent": Unknown word.
7 7,2019,CS,30,Huawei,Mon9r "Huawei": Unknown word.
8 8,2019,CS,31,Tiger,Mon10r
9 9,2019,CS,30,Scot,Mon11r "Scot": Unknown word.
10 10,2019,CS,31,Xujun Jiang,Mon12r "Xujun": Unknown word.
11 |
```

图 1: 生成的班级信息文件，命名为 ClassInfoTable.txt



```
TeX Report.tex 1, M | ClassInfoTable.txt U X | StudentScoreTable.txt U X
DatabaseSystem > Experiments > Experiment_04 > Files > StudentScoreTable.txt
1 114514 SunistC 201903 202001 DatabaseSystem 10086
2 114515 Jecosine 201904 202001 DatabaseSystem 1919810 "Jecosine"
3 114516 sdfa 201906 202001 DatabaseSystem 11451 "sdfa": Unknown word.
4 114517 asdfasd 201908 202001 DatabaseSystem 14514 "asdfasd": Unknown word.
5 114518 weasdf 201910 202001 DatabaseSystem 19810 "weasdf": Unknown word.
6 114519 werqwe 201912 202001 DatabaseSystem 23333 "werqwe": Unknown word.
7 114520 asdfzxcv 201914 202001 DatabaseSystem 314159 "asdfzxcv"
8 114521 asdfaxcvc 201916 202001 DatabaseSystem 271828 "asdfaxcvc"
9 114522 sadrqertr 201918 202001 DatabaseSystem 65535 "sadrqertr"
10 114523 dfgfdh 201920 202001 DatabaseSystem 255 "dfgfdh": Unknown word.
11 |
```

图 2: 生成的学生成绩信息文件，命名为 StudentScoreTable.txt

然后我们将其添加到我们的数据库中，其中，班级信息表我们选用"Comma-separated"模式导入，学生成绩信息表我们选用"Tab-separated"模式导入，如下图 3 和图 4 所示，导入结果如下图 5 和图 6 所示。

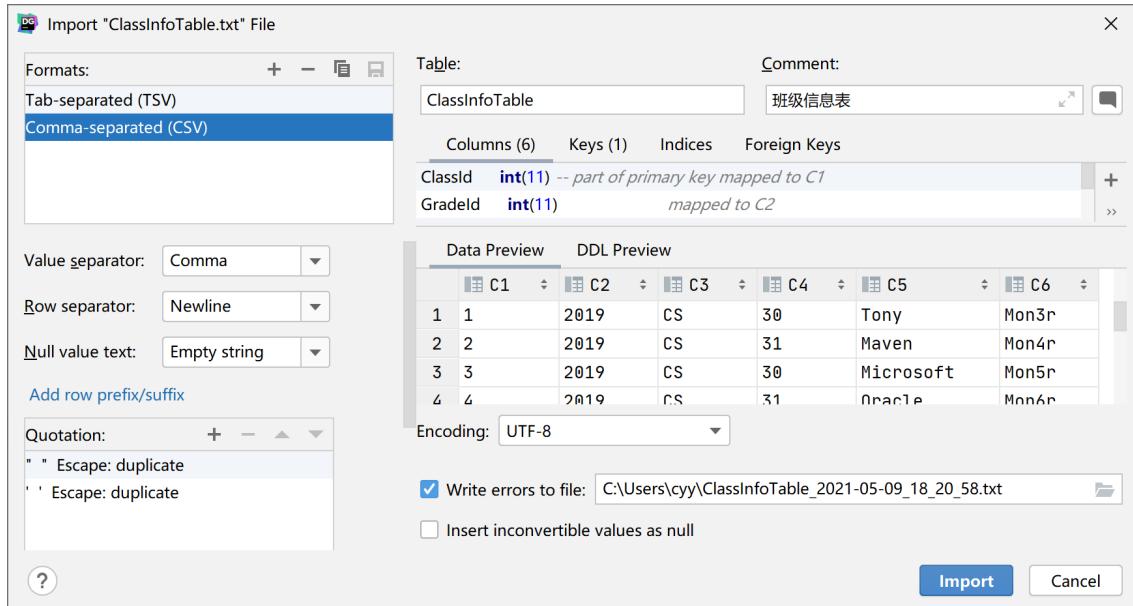


图 3: 使用"Comma-separated" 模式导入班级信息表

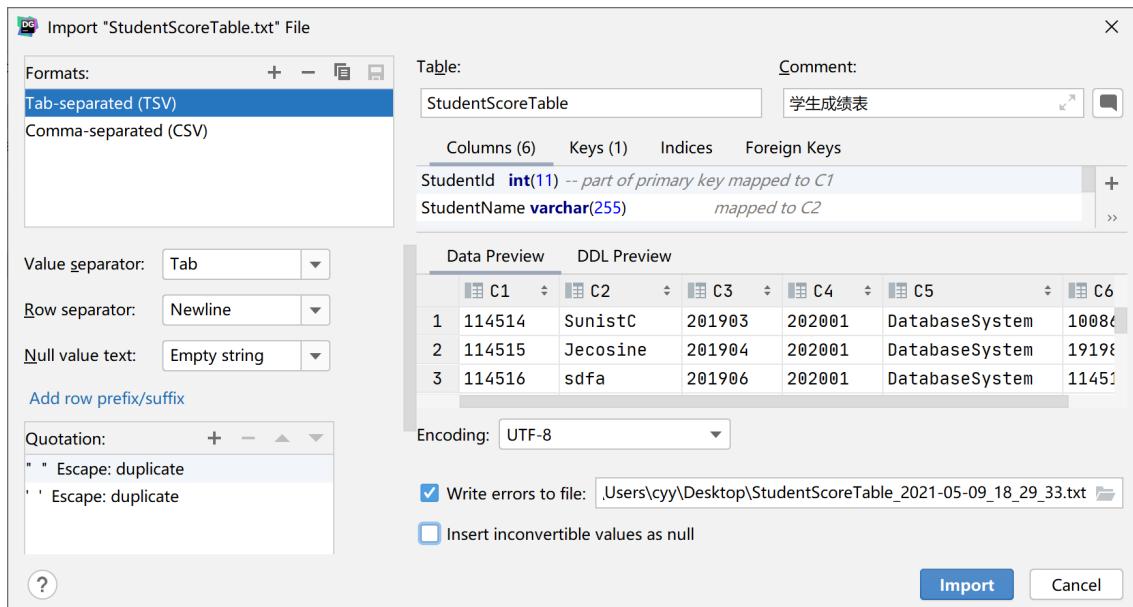


图 4: 使用"Tab-separated" 模式导入学生成绩信息表

The screenshot shows the DataGrip IDE interface. The top menu includes File, Edit, View, Navigate, Code, Refactor, Run, Tools, VCS, Window, Help, and a tab for 'default - console'. The Database section shows 'TencentCloud-CynosDB-N1' selected. The Services section shows 'TencentCloud-Cynos' and its services. The main area displays the results of a 'select * from ClassInfoTable' query. The results grid has columns: ClassId, GradeId, MajorName, StudentCount, TeacherName, and MasterName. The data shows 10 rows of student information.

ClassId	GradeId	MajorName	StudentCount	TeacherName	MasterName
1	1	2019 CS	30	Tony	Mon3r
2	2	2019 CS	31	Maven	Mon4r
3	3	2019 CS	30	Microsoft	Mon5r
4	4	2019 CS	31	Oracle	Mon6r
5	5	2019 CS	30	Bytedance	Mon7r
6	6	2019 CS	31	Tencent	Mon8r
7	7	2019 CS	30	Huawei	Mon9r
8	8	2019 CS	31	Tiger	Mon10r
9	9	2019 CS	30	Sccot	Mon11r
10	10	2019 CS	31	Xujun Jiang	Mon12r

图 5: 导入完成的班级信息表

The screenshot shows the DataGrip IDE interface. The top menu includes File, Edit, View, Navigate, Code, Refactor, Run, Tools, VCS, Window, Help, and a tab for 'default - console'. The Database section shows 'TencentCloud-CynosDB-N1' selected. The Services section shows 'TencentCloud-Cynos' and its services. The main area displays the results of a 'select * from StudentScoreTable' query. The results grid has columns: StudentId, StudentName, ClassId, TermId, CourseName, and Score. The data shows 10 rows of student scores.

StudentId	StudentName	ClassId	TermId	CourseName	Score
114514	SunistC	201903	202001	DatabaseSystem	10086
114515	Jecosine	201904	202001	DatabaseSystem	1919810
114516	sdfa	201906	202001	DatabaseSystem	11451
114517	asdfasd	201908	202001	DatabaseSystem	14514
114518	weasdf	201910	202001	DatabaseSystem	19810
114519	werqwe	201912	202001	DatabaseSystem	23333
114520	asdfzxcv	201914	202001	DatabaseSystem	314159
114521	asdfaxvcv	201916	202001	DatabaseSystem	271828
114522	sadrgertr	201918	202001	DatabaseSystem	65535
114523	dfgfdh	201920	202001	DatabaseSystem	255

图 6: 导入完成的学生成绩信息表

3.1.2 视图方式添加数据

在视图界面中，向课程信息表的添加 10 条数据，自己各构造。

我们构造了十条课程信息表的数据在 DataGrip 的视图界面内，如下图 7 所示。

The screenshot shows the DataGrip interface with the 'CourseInfoTable' view selected. The table contains 10 rows of data:

CourseId	CourseName	TeacherName
1	DatabaseSystem	XujunJiang
2	DataStructures	HailingXiong
3	OperatingSystem	HuafengDing
4	Algorithm	XingjiangXiao
5	AdvancedMath	LiyanJiang
6	DigitalImageProces...	Myself
7	DesignPatterns	Myself
8	IntroductionOfComp...	JikuanHu
9	SoftwareEngineering	Unknown
10	ComputerNetwork	JikuanHu

The Services panel shows two recent transactions: one for inserting data into the 'CourseInfoTable' and another for selecting all rows from it.

图 7: 构造的课程信息表

然后我们执行添加指令 (Ctrl+Enter)，结果如下图 8 所示。

The screenshot shows the DataGrip interface with the 'CourseInfoTable' view selected. The table now displays 10 rows of data, identical to those in Figure 7.

The Services panel shows the transaction history, including the successful insertion of data into the 'CourseInfoTable'.

图 8: 导入完成的课程信息表

3.1.3 导出数据表信息

在视图界面中，导出数据表信息，包含 Excel 格式和 TXT 文件格式。

我们在 DataGrip 中进行导出操作，过程如下图 9 所示，由于 TSV,CSV,XLSX 格式差别只有导出格式，此处省略了其余格式步骤。

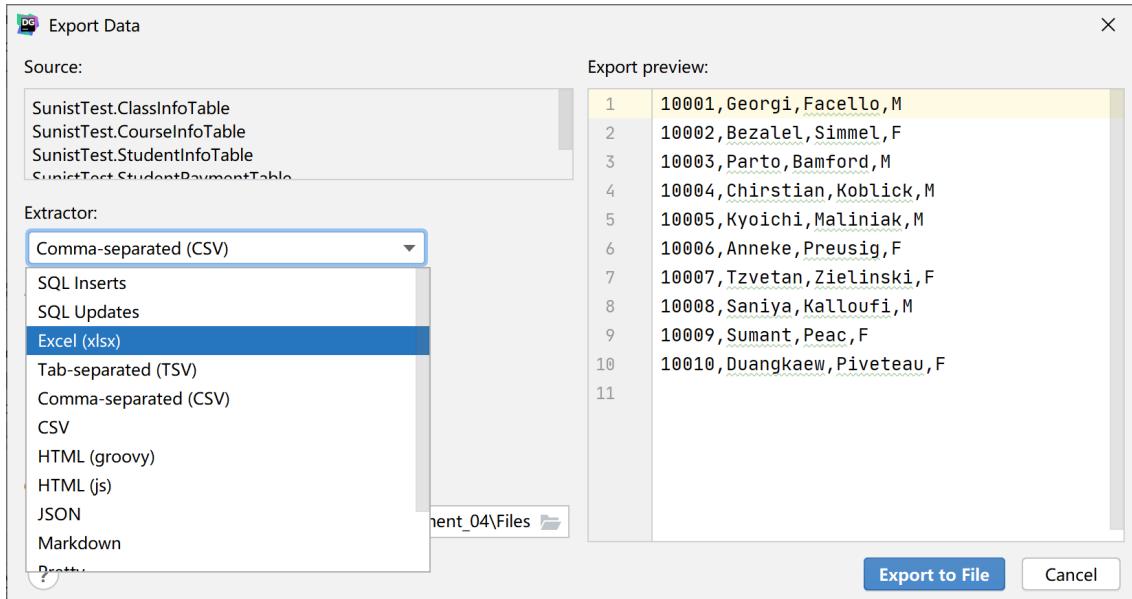


图 9: 导出数据库

导出完毕的数据库(此处列举三张表)如下：

1. CommaSeparatedVersion(CSV 格式)
2. TabSeparatedVersion(TSV 格式)
3. Excel 工作表 (XLSX 格式)

其中，CSV 格式使用逗号分隔，导出的是班级信息表；TSV 格式使用的是制表符分隔，导出的是班级信息表；XLSX 是 Excel 工作表格式，导出的是学生成绩信息表。

其中 CSV 能被作者的 Visual Studio Code 插件读取，所以显示为表格而不是文本文件。三张表格如下图 10 - 12 所示：

TEX Report.tex 1, M | SunistTest_ClassInfoTable.csv (read-only) U X

DatabaseSystem > Experiments > Experiment_04 > Files > SunistTest_ClassInfoTable.csv

	A	B	C	D	E	F
1	1	2019	CS	30	Tony	Mon3r
2	2	2019	CS	31	Maven	Mon4r
3	3	2019	CS	30	Microsoft	Mon5r
4	4	2019	CS	31	Oracle	Mon6r
5	5	2019	CS	30	Bytedance	Mon7r
6	6	2019	CS	31	Tencent	Mon8r
7	7	2019	CS	30	Huawei	Mon9r
8	8	2019	CS	31	Tiger	Mon10r
9	9	2019	CS	30	Sccot	Mon11r
10	10	2019	CS	31	Xujun Jiang	Mon12r

图 10: 导出完毕的 CSV 格式数据

TEX Report.tex 1, M | SunistTest_ClassInfoTable.tsv U X

DatabaseSystem > Experiments > Experiment_04 > Files > SunistTe

```

1 1 2019 CS 30 Tony Mon3r
2 2 2019 CS 31 Maven Mon4r
3 3 2019 CS 30 Microsoft Mon5r
4 4 2019 CS 31 Oracle Mon6r
5 5 2019 CS 30 Bytedance Mon7r
6 6 2019 CS 31 Tencent Mon8r "Ten
7 7 2019 CS 30 Huawei Mon9r "Hu
8 8 2019 CS 31 Tiger Mon10r
9 9 2019 CS 30 Sccot Mon11r "Sc
10 10 2019 CS 31 Xujun Jiang Mon12r
..
```

图 11: 导出完毕的 TSV 格式数据

TEX Report.tex 1, M | SunistTest_StudentScoreTable.xlsx U X

DatabaseSystem > Experiments > Experiment_04 > Files > SunistTest_StudentScoreTable.xlsx

	A	B	C	D	E	F
1	StudentId	StudentName	ClassId	TermId	CourseName	Score
2	114514	SunistC	201903	202001	DatabaseSystem	10086
3	114515	Jecosine	201904	202001	DatabaseSystem	1919810
4	114516	sdfa	201906	202001	DatabaseSystem	11451
5	114517	asdfasd	201908	202001	DatabaseSystem	14514
6	114518	weasdf	201910	202001	DatabaseSystem	19810
7	114519	werqwe	201912	202001	DatabaseSystem	23333
8	114520	asdfzxcv	201914	202001	DatabaseSystem	314159
9	114521	asdfaxcvc	201916	202001	DatabaseSystem	271828
10	114522	sadrqertr	201918	202001	DatabaseSystem	65535
11	114523	dfgfdh	201920	202001	DatabaseSystem	255

图 12: 导出完毕的 XLSX 格式数据

3.2 命令方式建立数据表

向实验三教材第三章的例子用到的三张表添加数据²。

3.2.1 使用 Insert 命令添加数据

用 Insert 命令向表（Course）中插入书上教材第三章例子中的数据。

输入的指令如下图 13 所示。

The screenshot shows the DataGrip interface with two tabs: 'Report.tex' and 'Report.sql'. The 'Report.sql' tab is active, displaying the following SQL code:

```
TEX Report.tex 1, M • | Report.sql M • |
DatabaseSystem > Experiments > Experiment_04 > Report > Report.sql
Run on active connection | Select block
# 使用Insert命令添加数据
CREATE TABLE IF NOT EXISTS Course (
    `Cno` INT,
    `Cname` VARCHAR(8) CHARACTER SET utf8,
    `Ccredit` INT,
    `Cpno` INT
);
INSERT INTO Course
VALUES (1, '数据库', 4, 5),
       (2, '数学', 6, NULL),
       (3, '信息系统', 3, 1),
       (4, '操作系统', 4, 6),
       (5, '数据结构', 4, 7),
       (6, '数据处理', 3, NULL),
       (7, 'PASCAL语言', 4, 6),
       (8, '计算机基础', 5, NULL),
       (9, 'C++程序设计', 4, 8),
       (10, '离散数学', 4, NULL),
       (11, '微机原理', 5, 8),
       (12, '数据库原理', 4, 9),
       (13, '计算机网络', 3, 9),
       (14, '软件工程', 3, 9),
       (15, '图形学', 3, 8);
```

图 13: 使用 Insert 命令添加数据的指令

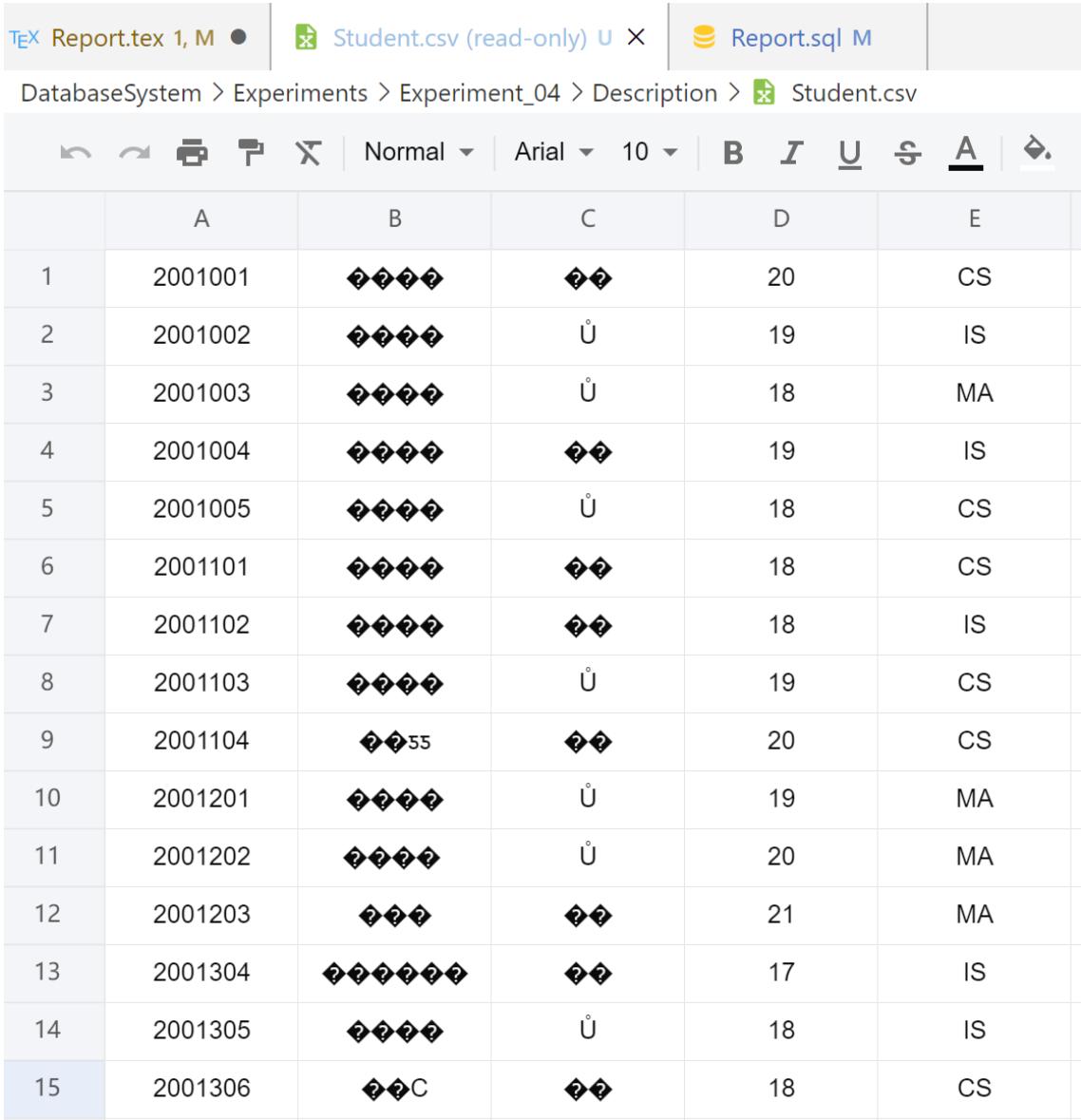
3.2.2 使用命令导入 Excel 工作表中的数据

此处我们在查询了 Stackoverflow, CSDN, GitHub, Microsoft 论坛, Oracle 论坛后, 未发现 MySQL 能够直接导入 xls/xlsx 格式的数据³, 必须使用中间件、第三方库或对 MySQL 进行二次开发等方式才能够导入 xls/xlsx 格式的数据, 此处我们将其保存为 csv 格式的数据, 进行导入。

²此处作者使用了 DataGrip 的控制台与 MysqClient 结合进行命令操作

³不能导入的原因在总结处

我们先将实验给定的 xls 格式 Student 表导出为 csv 表格。导出结果如下图 14 所示。



The screenshot shows a CSV file named 'Student.csv' in a spreadsheet application. The file contains 15 rows of data with columns A through E. The data includes various characters such as numbers, letters, and special characters like 'Ü'. The first few rows show the following data:

	A	B	C	D	E
1	2001001	?????	??	20	CS
2	2001002	?????	Ü	19	IS
3	2001003	?????	Ü	18	MA
4	2001004	?????	??	19	IS
5	2001005	?????	Ü	18	CS
6	2001101	?????	??	18	CS
7	2001102	?????	??	18	IS
8	2001103	?????	Ü	19	CS
9	2001104	??55	??	20	CS
10	2001201	?????	Ü	19	MA
11	2001202	?????	Ü	20	MA
12	2001203	???	??	21	MA
13	2001304	???????	??	17	IS
14	2001305	?????	Ü	18	IS
15	2001306	??C	??	18	CS

图 14: 导出 Student 为 CSV 格式

为什么会出现乱码呢?是实验给定的表格中文字集使用的是 GBK 编码(Windows Notepad 的编码方式), 而作者的预览器使用的是与 MySQL 一致的 UTF-8 格式, 两者并不兼容导致的。

在作者经过一晚上四个小时的探索后, 发现本实验要求的使用命令"LOAD DATA LOCAL INFILE {\$filename} INTO TABLE {\$tablename}" 导入数据时, DataGrip 所要求 MySQL 最低版本为⁴⁸, 错误信息如下图 15 所示, 所以作者只能使用 MysqlClient 登录进行操作。

由于 csv 数据使用的字符编码是 GBK, 而数据库的编码为 UTF8, 所以导入会出错, 如下图 16 所示, 我们改用 GBK 作为字符编码导入, 如下图 16 所示。

导出后的结果如下图 17 所示, 因为作者先前强制使用 UTF8 导入了 GBK 编码的字符, 所以前 15 条记录是乱码记录。

⁴此处作者使用的 MySQL 版本为 5.7.12(腾讯云分布式数据库定制版)

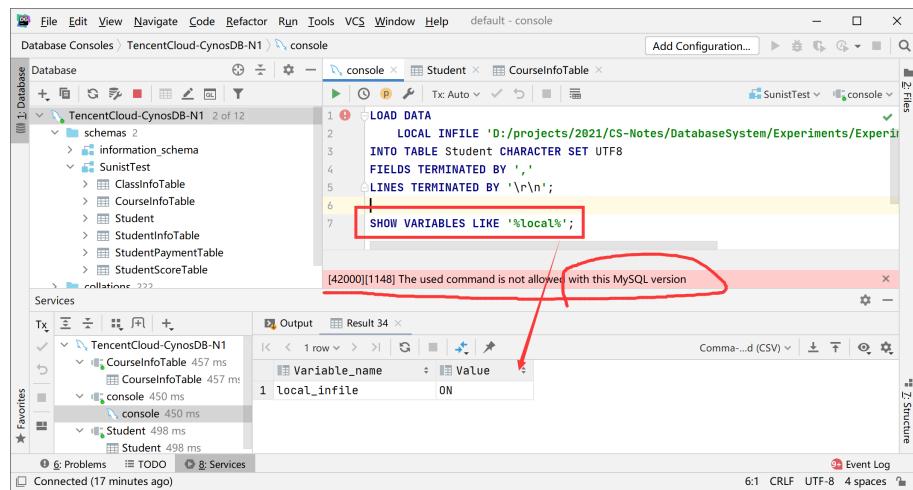


图 15: DataGrip 不允许此版本的 MySQL 使用 LOAD DATA INFILE 指令

```

mysql> LOAD DATA
    -> LOCAL INFILE 'D:/projects/2021/CS-Notes/DatabaseSystem/Experiments/Experiment_04/Description/Student.csv'
    -> INTO TABLE Student CHARACTER SET UTF8
    -> FIELDS TERMINATED BY ','
    -> LINES TERMINATED BY '\r\n';
ERROR 1300 (HY000): Invalid utf8 character string: ''
mysql> LOAD DATA
    -> LOCAL INFILE 'D:/projects/2021/CS-Notes/DatabaseSystem/Experiments/Experiment_04/Description/Student.csv'
    -> INTO TABLE Student CHARACTER SET GBK
    -> FIELDS TERMINATED BY ','
    -> LINES TERMINATED BY '\r\n';
Query OK, 15 rows affected (0.07 sec)
Records: 15 Deleted: 0 Skipped: 0 Warnings: 0
mysql> |

```

图 16: 使用 MysqClient 进行导入操作

The screenshot shows the DataGrip interface with the 'Student' table selected. The table has the following columns: Sno, Sname, Ssex, Sage, and Sdept. The data is as follows:

	Sno	Sname	Ssex	Sage	Sdept
1	2001001	李勇	男	20	CS
2	2001002	刘晨	女	19	IS
3	2001003	王敏	女	18	MA
4	2001004	张立	男	19	IS
5	2001005	刘云	女	18	CS
6	2001181	陈林	男	18	CS
7	2001182	程明	男	18	IS
8	2001183	王艳	女	19	CS
9	2001184	严平平	男	20	CS
10	2001281	王敏	女	19	MA
11	2001282	孙祥欣	男	20	MA
12	2001283	孙研	女	21	MA
13	2001304	孙小	男	17	IS
14	2001305	孙林	女	18	IS
15	2001306	李计	男	18	CS
16	2001001	李勇	男	20	CS
17	2001002	刘晨	女	19	IS
18	2001003	王敏	女	18	MA
19	2001004	张立	男	19	IS
20	2001005	刘云	女	18	CS
21	2001181	陈林	男	18	CS
22	2001182	程明	男	18	IS
23	2001183	王艳	女	19	CS
24	2001184	严平平	男	20	CS
25	2001281	王敏	女	19	MA
26	2001282	孙祥欣	男	20	MA
27	2001283	孙研	女	21	MA
28	2001304	孙林	男	17	IS
29	2001305	李计	女	18	IS
30	2001306	孙小	男	18	CS

图 17: 导入完成的结果

3.3 使用命令导入 txt 格式的数据

由于 csv 格式其实就是 txt 格式，上节已进行过，本节略。

3.4 使用命令导出数据表

我们使用 SELECT INTO OUTFILE 指令进行导出，其指令如下图 18 所示。

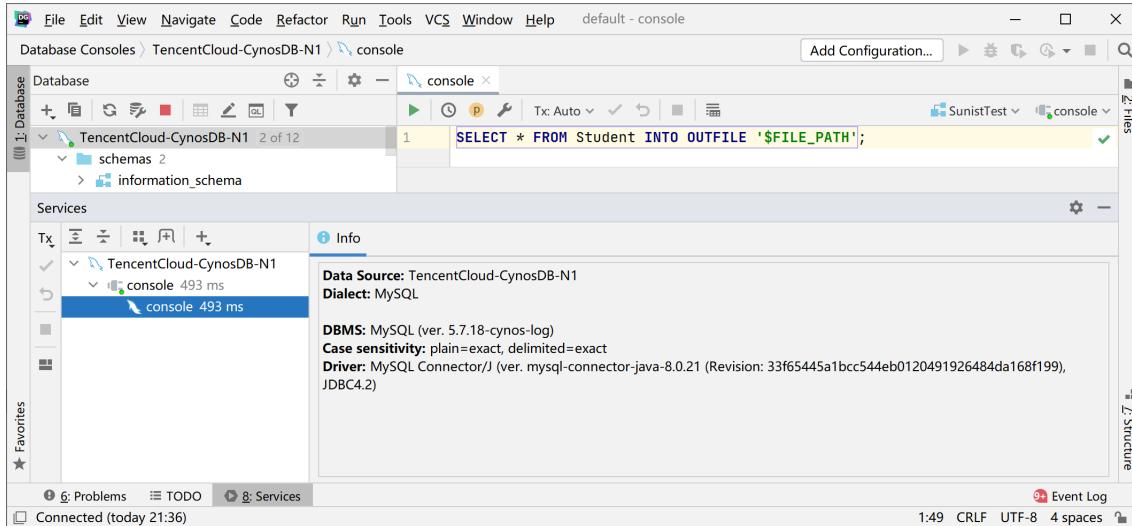


图 18: 使用 Select Into Outfile 导出的过程

但由于此处使用的是远程服务器，作者没有读取服务器文件的权限⁵，所以此处作者只能使用 mysqldump 进行导出，导出的文件为.sql 文件，如下图 19 所示。

```
cyy@Sunist-WorkStation ~ ✘ USAGE
> mysqldump -u admin -p -P 24307 -h gz-cynosdbmysql-grp-cj7ardbz.sql.tencentcdb.com --default-character-set=utf8 Sunistest > ~\Desktop\dmp.sql
Enter password: *****
Warning: A partial dump from a server that has GTIDs will by default include the GTIDs of all transactions, even those that changed suppressed parts of the database. If you don't want to restore GTIDs, pass --set-gtid-purged=OFF. To make a complete dump, pass --all-databases --triggers --routines --events.
mysqldump: Couldn't execute 'SELECT COLUMN_NAME, JSON_EXTRACT(HISTOGRAM, ".$."number-of-buckets-specified") FROM information_schema.COLUMN_STATISTICS WHERE SCHEMA_NAME = 'SunistTest' AND TABLE_NAME = 'ClassInfoTable';': Unknown table 'COLUMN_STATISTICS' in information_schema (1109)
cyy@Sunist-WorkStation ~ ✘ USAGE
```

图 19: 使用 Mysqldump 导出的过程

导出结果如下图 20 所示 (此处导出的是表结构)，在本地数据库重新建表，执行普通导出操作即可⁶。

⁵ 此处作者只能使用 Mysql 的端口与 Mysql 进行通信，而无法通过 Mysql 读取其导出的文件，当然远程服务器也没有给予 Mysql 可以导出文件的权限

⁶ 此处作者本地未安装有 Mysql-Server，故无法进行此实验

图 20: 使用 Mysqldump 导出的结果

4 实验总结

4.1 Excel 无法导入的原因

因为 MySQL 的 load data infile 是通过字符流的方式进行读写操作的。而 Excel 内部的结构根本不是文件，而是一个 xml 文件作为工作簿的信息，加上三个文件夹 _rels, xl 和 docProps 存储数据信息的，其中工作表信息存储在 xl 文件夹中，内部真实结构是 xml 标记文档，虽然其本质依旧是文本文件，但是其信息已经被复杂的文件层级结构加上 Office 的标记给层层覆盖，使得 MySQL 无法直接进行读取。读取此类文件，需要先将xlsx文件解压缩，再提取其表格数据文件，剥离 Office 元余标记后才能进行读取，而 MySQL 的 load data infile 指令根本无法进行此类复杂工作，必须借助第三方工具。其中，xlsx 的层级结构如下图 21 所示，数据表的数据信息如下图 22 所示。

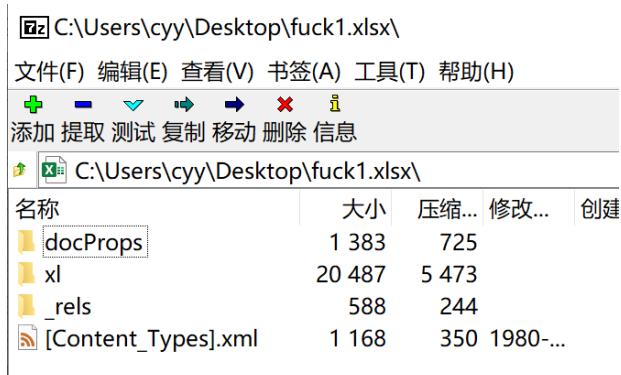


图 21: xlsx 的内部文件结构

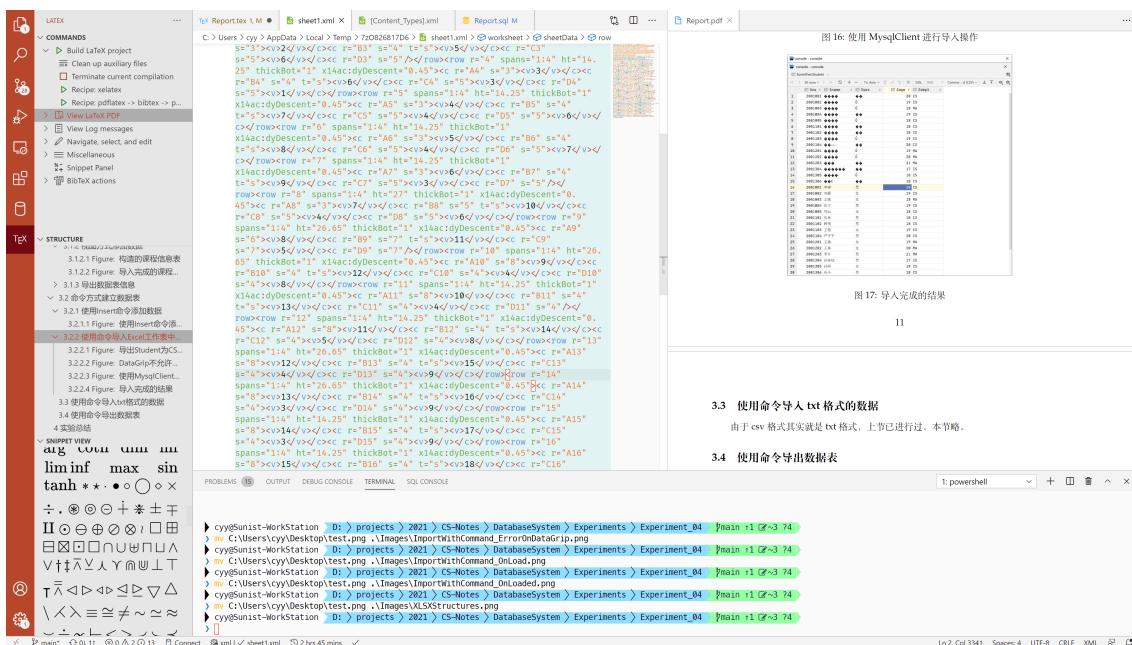


图 22: xlsx 存储工作表的内部数据

4.2 心得

1. 设计数据表时一定要注意字符集，此项一定要与产生数据的后端、前端沟通好，避免因为字符集的问题出错。目前互联网普遍使用的字符集为 utf8mb4，此字符集不仅能存储所有的常见字，也能存储诸如 emoji 表情一类的拓展文字，但所使用的数据库驱动引擎必须为 InnoDB。
2. 要善于从 Stackoverflow, GitHub 等论坛寻找解决问题的方法，如此次实验设置了无法完成的操作，不经过深入探讨根本无法得知出错的原因。
3. 使用 DataGrip, WorkBench, Navicat 等图形化数据库管理系统时，需要先了解其支持指令、所需数据库版本等信息，此次实验出现的‘used command is not allow with this MySQL Version’等问题是应该要事前知道并寻找替代解决方案的。
4. L^AT_EX 的字体在 macOS 与 Windows 并不一样，如在 mac/Linux 上显示为‘FiraCode NF’的等宽字体在 Windows 上显示为‘FiraCode Nerd Font’(同一个 TTF 文件)，L^AT_EX 是无法编译的。