《数据管理基础》Lab2 实验报告

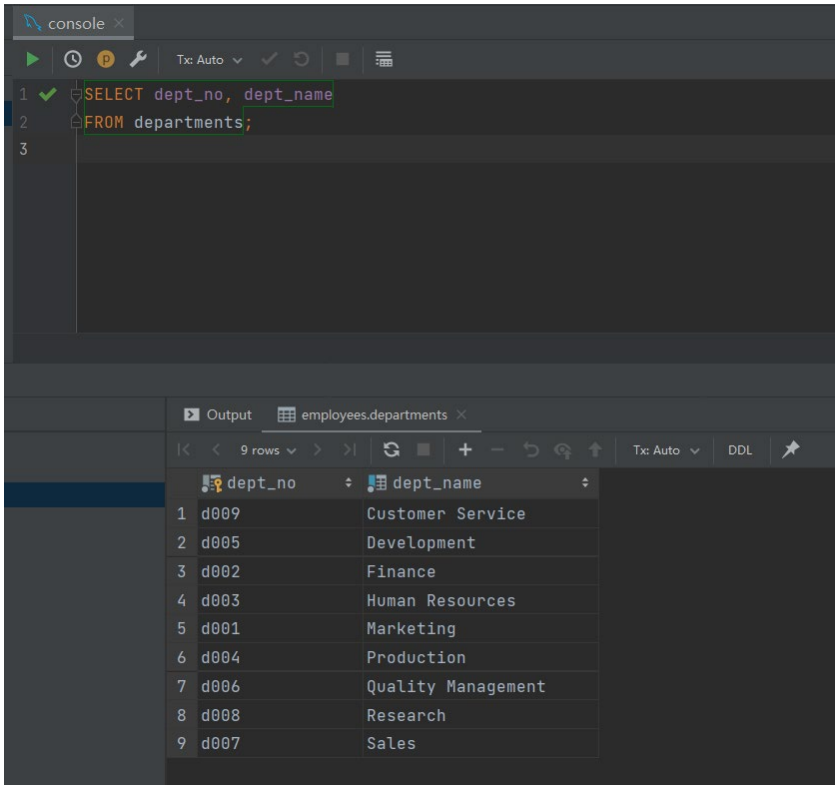
姓名 郁博文 学号 201250070 联系方式 201250070@smail.nju.edu.cn

一、实验环境

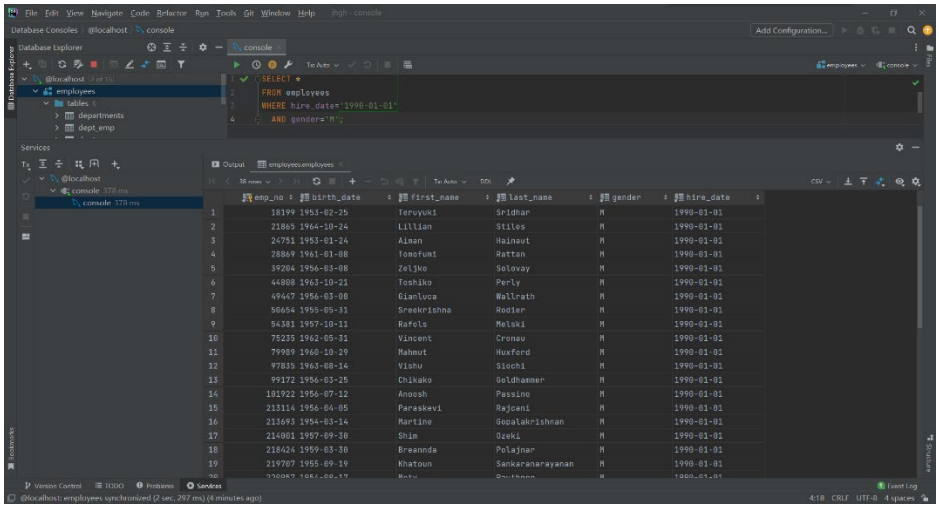
Windows10;
MySQL Shell 8.0.27;
MySQL Workbench 8.0 CE;
JetBrains DataGrip 2021.3

二、实验过程

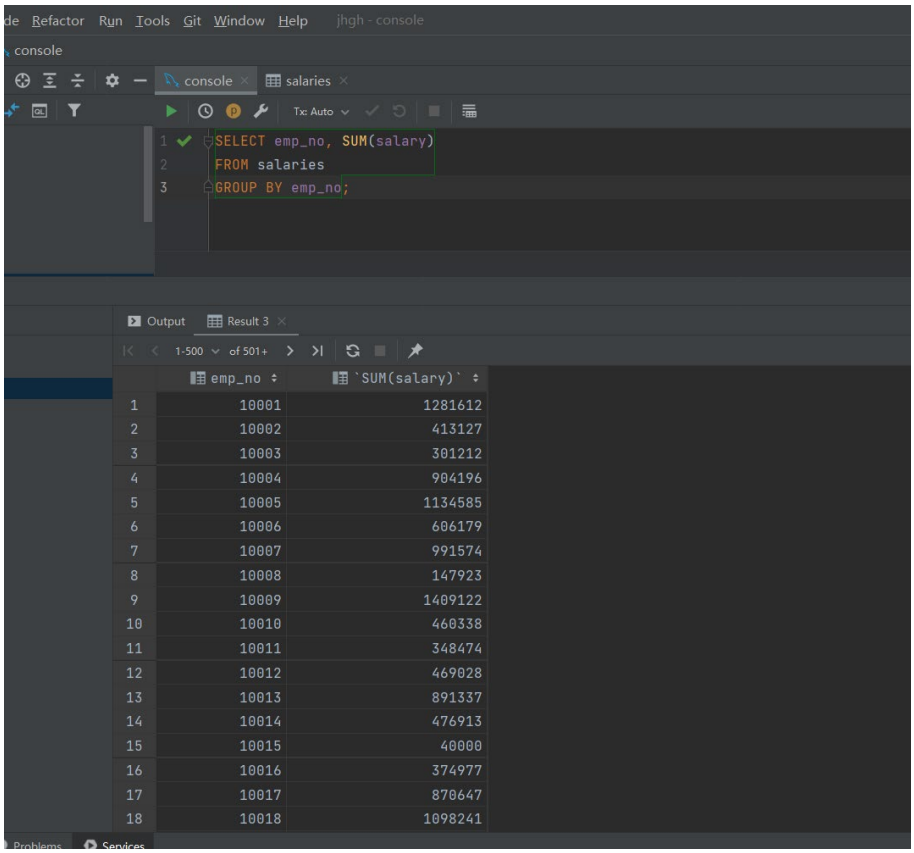
1、单表查询（查询）



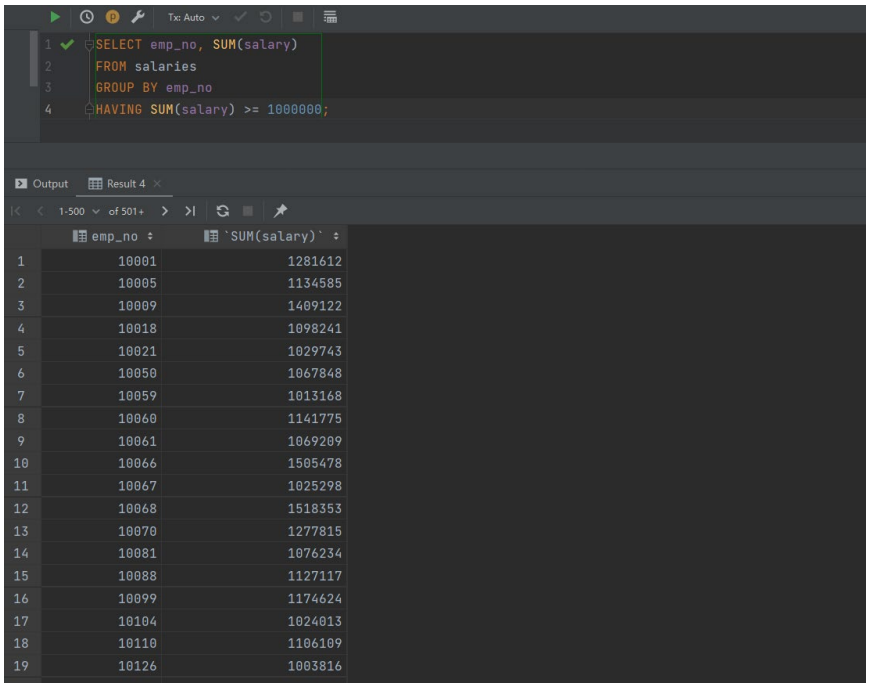
2、单表查询（选择）



3. 不带分组过滤条件的分组统计查询



4. 带分组过滤条件的分组统计查询



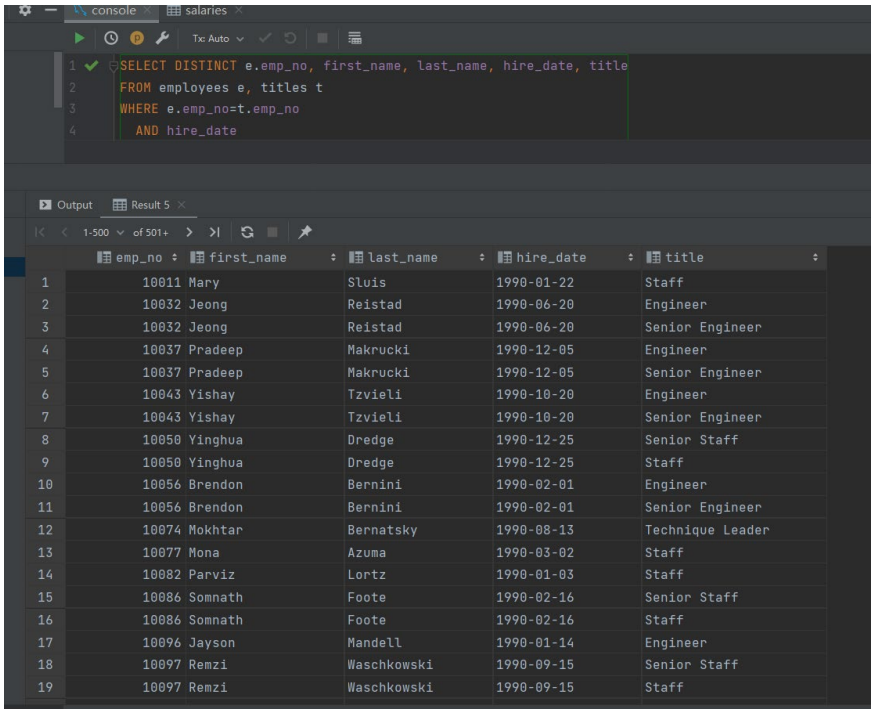
```
SELECT emp_no, SUM(salary)
FROM salaries
GROUP BY emp_no
HAVING SUM(salary) >= 1000000;
```

Output Result 4

1-500 of 501+

	emp_no	`SUM(salary)`
1	10001	1281612
2	10005	1134585
3	10009	1409122
4	10018	1098241
5	10021	1029743
6	10050	1067848
7	10059	1013168
8	10060	1141775
9	10061	1069209
10	10066	1505478
11	10067	1025298
12	10068	1518353
13	10070	1277815
14	10081	1076234
15	10088	1127117
16	10099	1174624
17	10104	1024013
18	10110	1106109
19	10126	1003816

5. 两表连接查询（普通连接）



```
SELECT DISTINCT e.emp_no, first_name, last_name, hire_date, title
FROM employees e, titles t
WHERE e.emp_no=t.emp_no
  AND hire_date
```

Output Result 5

1-500 of 501+

	emp_no	first_name	last_name	hire_date	title
1	10011	Mary	Sluis	1990-01-22	Staff
2	10032	Jeong	Reistad	1990-06-20	Engineer
3	10032	Jeong	Reistad	1990-06-20	Senior Engineer
4	10037	Pradeep	Makrucki	1990-12-05	Engineer
5	10037	Pradeep	Makrucki	1990-12-05	Senior Engineer
6	10043	Yishay	Tzvieli	1990-10-20	Engineer
7	10043	Yishay	Tzvieli	1990-10-20	Senior Engineer
8	10050	Yinghua	Dredge	1990-12-25	Senior Staff
9	10050	Yinghua	Dredge	1990-12-25	Staff
10	10056	Brendon	Bernini	1990-02-01	Engineer
11	10056	Brendon	Bernini	1990-02-01	Senior Engineer
12	10074	Mokhtar	Bernatsky	1990-08-13	Technique Leader
13	10077	Mona	Azuma	1990-03-02	Staff
14	10082	Parviz	Lortz	1990-01-03	Staff
15	10086	Somnath	Foote	1990-02-16	Senior Staff
16	10086	Somnath	Foote	1990-02-16	Staff
17	10096	Jayson	Mandell	1990-01-14	Engineer
18	10097	Remzi	Waschkowski	1990-09-15	Senior Staff
19	10097	Remzi	Waschkowski	1990-09-15	Staff

6.

Q1：不在 GROUP BY 子句中出现的属性，是否可以出现在 SELECT 子句中？请举例并上机验证。

任何没有出现在 groupby 子句中的属性只能出现在聚集函数内部，否则就是错误的。

```
SELECT emp_no,salary,SUM(salary)
FROM salaries
GROUP BY emp_no;
```

Result 6

1-500 of 501+

emp_no	salary	`SUM(salary)`
10001	60117	1281612
10002	65828	413127
10003	40006	301212
10004	40054	904196
10005	78228	1134585
10006	40000	606179
10007	56724	991574
10008	46671	147923
10009	60929	1409122
10010	72488	460338
10011	42365	348474
10012	40000	469028
10013	40000	891337
10014	46168	476913

WHERE ORDER BY

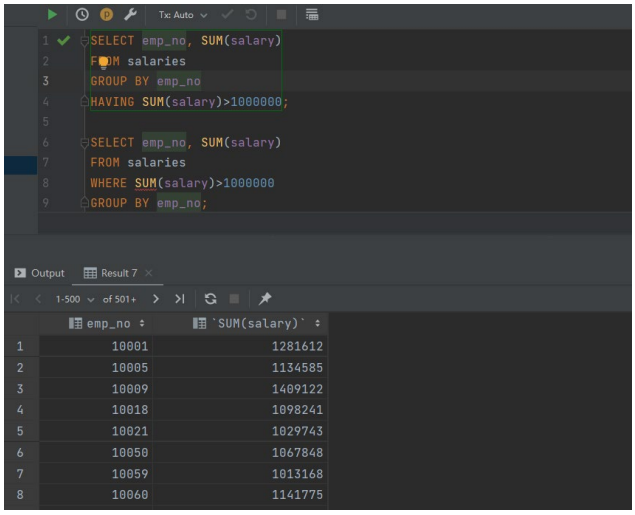
	emp_no	salary	from_date	to_date
1	10001	60117	1986-06-26	1987-06-26
2	10001	62102	1987-06-26	1988-06-25
3	10001	66074	1988-06-25	1989-06-25
4	10001	66596	1989-06-25	1990-06-25
5	10001	66961	1990-06-25	1991-06-25
6	10001	71046	1991-06-25	1992-06-24
7	10001	74333	1992-06-24	1993-06-24

如图所示，执行上述错误 SQL 查询语句后，emp_no 为 10001 的记录不止一条，因此对于这样的分组结果，salary 属性显示为 salaries 表中 emp_no=10001 的记录的第一个 salary 值，而这是没有意义的。

Q2：请举例说明分组统计查询中 WHERE 和 HAVING 有何区别？

WHERE 子句作用于基本表或者视图，而 HAVING 作用于分组结果。

同时，HAVING 子句中可以出现聚集函数而 WHERE 子句不可以。



可以看到 DataGrip 直接指出了第二条查询语句中 WHERE 子句的错误。

Q3：连接查询速度是影响关系数据库性能的关键因素。请讨论如何提高连接查询速度，并进行实验验证。

由于连接算法不同，因此连接操作的速度也不一样。循环嵌套算法是最慢的。如果连接条件只有等值比较，则容易优化（有序双指针比较）；如果连接条件能够建立索引，且事先已建立索引，则容易优化；

```
employees> SELECT e.emp_no, d.dept_name, e.hire_date
           FROM employees e, dept_emp de, departments d
           WHERE e.emp_no=de.emp_no
           AND de.dept_no=d.dept_no
           AND e.hire_date='1992-01-01'
[2022-03-04 19:41:37] 71 rows retrieved starting from 1 in 138 ms (execution: 117 ms, fetching: 21 ms)
employees> SELECT COUNT(*) FROM (
             SELECT e.emp_no, d.dept_name, e.hire_date
             FROM employees e, dept_emp de, departments d
             WHERE e.emp_no=de.emp_no
             AND de.dept_no=d.dept_no
             AND d.dept_name='Finance'
           ) AS C
[2022-03-04 19:41:37] 1 row retrieved starting from 1 in 264 ms (execution: 118 ms, fetching: 146 ms)
```

三、实验中遇到的问题及解决办法

本次实验无。

四、参考文献及致谢

实验手册：https://www.programminghunter.com/article/17522038636/