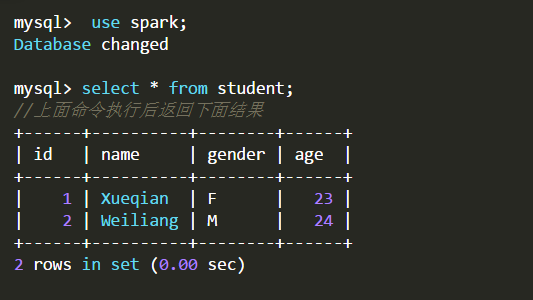
**作业8**

**使用Spark SQL读写MySQL数据库**

1. **在MySQL数据库中创建表**

在MySQL数据库中创建一个名称为spark的数据库，并创建一个名称为student的表，录入两条数据，如下：



1. **在spark-shell交互式执行环境中读写MySQL数据库**

继续上一步的操作。

（1）在spark-shell交互式执行环境中，撰写Scala代码，读取MySQL数据库的内容；

（2）在spark-shell交互式执行环境中，撰写Scala代码，向MySQL数据库写入两条新的记录：

|  |
| --- |
| | 3 | Rongcheng | M | 26 |  | 4 | Guanhua | M | 27 | |

1. **在IDEA中开发程序MySQL数据库读写数据**

继续上一步的操作。

（1）在IDEA中创建Maven工程，开发Spark SQL程序，从MySQL数据库读取数据。

（2）在IDEA中创建Maven工程，开发Spark SQL程序，向MySQL数据库写入两条新的记录：

|  |
| --- |
| | 5 | Chenglu | F | 22 |  | 6 | Linzhe | M | 23 | |

（备注：需要把Scala代码文件内容和pom.xml文件内容文本粘贴到实验报告正文中，不是截图，是文本格式，可以编辑）

|  |
| --- |
| 参考：  为Maven工程添加MySQL驱动等相关依赖：  <dependency>  <groupId>org.apache.spark</groupId>  <artifactId>spark-core\_2.11</artifactId>  <version>2.1.0</version>  <scope>provided</scope>  </dependency>  <dependency>  <groupId>org.apache.spark</groupId>  <artifactId>spark-sql\_2.11</artifactId>  <version>2.1.0</version>  <scope>provided</scope>  </dependency>  <dependency>  <groupId>mysql</groupId>  <artifactId>mysql-connector-java</artifactId>  <version>5.1.40</version>  </dependency> |

**四、实验报告**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 实验报告 | | | | |
| 题目 | 使用Spark SQL读写MySQL数据库 | 姓名 | 吴雨娟 | 日期：  2021.11.18 |
| 实验环境：VMware workstation 16 pro，Ubuntu-20.04操作系统, JDK1.8,mysql-8.0.27,scala-2.12.10, Spark-3.1.2,maven-3.8.3,ideaIU-2021.2.2 | | | | |
| 实验内容与完成情况：  一、在MySQL数据库中创建表  在MySQL数据库中创建一个名称为spark的数据库，并创建一个名称为student的表，录入两条数据。        二、在spark-shell交互式执行环境中读写MySQL数据库  继续上一步的操作。  （1）在spark-shell交互式执行环境中，撰写Scala代码，读取MySQL数据库的内容；      （2）在spark-shell交互式执行环境中，撰写Scala代码，向MySQL数据库写入两条新的记录：            三、在IDEA中开发程序MySQL数据库读写数据  继续上一步的操作。  （1）在IDEA中创建Maven工程，开发Spark SQL程序，从MySQL数据库读取数据。  1.pom.xml文件内容：  <?xml version="1.0" encoding="UTF-8"?>  <project xmlns="http://maven.apache.org/POM/4.0.0"  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">  <modelVersion>4.0.0</modelVersion>  <groupId>org.example</groupId>  <artifactId>wordcount3</artifactId>  <version>1.0-SNAPSHOT</version>  <properties>  <maven.compiler.source>8</maven.compiler.source>  <maven.compiler.target>8</maven.compiler.target>  </properties>  <dependencies>  <dependency>  <groupId>org.apache.spark</groupId>  <artifactId>spark-core\_2.12</artifactId>  <version>3.1.2</version>  </dependency>  <dependency>  <groupId>org.apache.spark</groupId>  <artifactId>spark-sql\_2.12</artifactId>  <version>3.1.2</version>  </dependency>  <dependency>  <groupId>mysql</groupId>  <artifactId>mysql-connector-java</artifactId>  <version>8.0.27</version>  </dependency>  </dependencies>  </project>  2. Scala代码文件内容  import java.sql.{ Connection, DriverManager }  object ScalaJdbcConnectSelect extends App{  val url = "jdbc:mysql://localhost:3306/spark"  val driver = "com.mysql.jdbc.Driver"  val username = "root"  val password = "hadoop"  var connection: Connection = \_  try {  Class.forName(driver)  connection = DriverManager.getConnection(url, username, password)  val statement = connection.createStatement  val rs = statement.executeQuery("SELECT \* FROM student")  while (rs.next) {  val id =rs.getString("id")  val name = rs.getString("name")  val gender = rs.getString("gender")  val age =rs.getString("age")  println("| %s | %s | %s | %s |".format(id,name, gender,age))  }  //println("查询数据完成！")  }  }  3.运行结果    （2）在IDEA中创建Maven工程，开发Spark SQL程序，向MySQL数据库写入两条新的记录：  1.pom.xml文件内容与（1）中相同。  2. Scala代码文件内容  import java.sql.{ Connection, DriverManager}  object ScalaJdbcConnectSelect extends App{  val url = "jdbc:mysql://localhost:3306/spark"  val driver = "com.mysql.jdbc.Driver"  val username = "root"  val password = "hadoop"  var connection: Connection = \_  try {  Class.forName(driver)  connection = DriverManager.getConnection(url, username, password)  val statement = connection.createStatement  val rs2 = statement.executeUpdate("INSERT INTO `student` (`id`, `name`, `gender`, `age`) VALUES ('5', 'Chenglu','F','22')")  val rs3 = statement.executeUpdate("INSERT INTO `student` (`id`, `name`, `gender`, `age`) VALUES ('6', 'Linzhe','M','23')")  val rs4 = statement.executeQuery("SELECT \* FROM student")  while (rs4.next) {  val id =rs4.getString("id")  val name = rs4.getString("name")  val gender = rs4.getString("gender")  val age =rs4.getString("age")  println("| %s | %s | %s | %s |".format(id,name, gender,age))  }  }  }  3.运行结果 | | | | |
| 出现的问题：  1.问题一：在IDEA中编写Scala程序向MySQL数据库写入两条新的记录时，写入后数据无法在IDEA中显示，但是在mysql中可以显示。 | | | | |
| 解决方案（列出遇到的问题和解决办法，列出没有解决的问题）：  1.问题一的解决方案：发现是Scala代码出现问题，经过排查，发现读取的数据不是最新的数据，所以显示不出来。把代码修改成添加新数据后，再读取所有数据，这样就能显示正确的结果。修改后部分如下图所示。 | | | | |

（备注：相关实验过程可以截图放入实验报告）