

3. Определите в какой нормальной форме данная таблица, приведите ее ко 2 и 3 нормальным формам последовательно.

```
File Edit Selection View Go Run ... hw1
hw1.scala x s1.scala
hw1.scala
1  /*
2  chcp 65001 && spark-shell -i \GeekBrains\Data_Engineer\ETL\hws\hw1\hw1.scala --conf "spark.driver.extraJavaOptions=-Dfile.encoding=utf-8"
3  */
4  import org.apache.spark.internal.Logging
5  import org.apache.spark.sql.functions.{col, collect_list, concat_ws}
6  import org.apache.spark.sql.{DataFrame, SparkSession}
7  val t1 = System.currentTimeMillis()
8  if(1==1){
9  var df1 = spark.read.format("com.crealytics.spark.excel")
10     .option("sheetName", "Sheet1")
11     .option("useHeader", "false")
12     .option("treatEmptyValuesAsNulls", "false")
13     .option("inferSchema", "true").option("addColorColumns", "true")
14     .option("usePlainNumberFormat", "true")
15     .option("startColumn", 0)
16     .option("endColumn", 99)
17     .option("timestampFormat", "MM-dd-yyyy HH:mm:ss")
18     .option("maxRowsInMemory", 20)
19     .option("excerptSize", 10)
20     .option("header", "true")
21     .format("excel")
22     .load("C:/GeekBrains/Data_Engineer/ETL/hws/hw1/hw1.xlsx")
23     df1.show()
24     df1.withColumn("id",monotonicallyIncreasingId)
25     .orderBy("id").drop("id","Name","Job", "Home_city")
26     .write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
27     .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tasketlhw1c")
28     .mode("overwrite").save()
29     df1.select(["Employee_ID","Name"]).distinct()
30     .write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
31     .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tasketlhw1d")
32     .mode("overwrite").save()
33 }
```

```
File Edit Selection View Go Run ... hw1
hw1.scala x s1.scala
hw1.scala
20 var df1 = spark.read.format("com.crealytics.spark.excel")
21     .option("header", "true")
22     .format("excel")
23     .load("C:/GeekBrains/Data_Engineer/ETL/hws/hw1/hw1.xlsx")
24     df1.show()
25     df1.withColumn("id",monotonicallyIncreasingId)
26     .orderBy("id").drop("id","Name","Job", "Home_city")
27     .write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
28     .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tasketlhw1c")
29     .mode("overwrite").save()
30     df1.select(["Employee_ID","Name"]).distinct()
31     .write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
32     .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tasketlhw1d")
33     .mode("overwrite").save()
34     df1.select("Job_Code","Job").distinct()
35     .write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
36     .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tasketlhw1e")
37     .mode("overwrite").save()
38     df1.select("City_code", "Home_city").distinct()
39     .write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
40     .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tasketlhw1f")
41     .mode("overwrite").save()
42     println("task 1")
43 }
44 val s0 = (System.currentTimeMillis() - t1)/1000
45 val s = s0 % 60
46 val m = (s0/60) % 60
47 val h = (s0/60/60) % 24
48 println("%02d:%02d:%02d".format(h, m, s))
49 System.exit(0)
```



Unamed-1\spark\taskethw1c\ - HeidiSQL 12.8.0.6908

Файл Редактировать Поиск Запрос Инструменты Переход Помощь

Фильтр баз даннь Фильтр таблиц

Unamed-1 База данных: spark Таблица: taskethw1c Данные Запрос

Unamed-1

- information\_schema
- mysql
- performance\_schema
- sakila
- spark 112,0 KiB
  - tasketh1a 16,0 KiB
  - tasketh1b 16,0 KiB
  - tasketh1b 16,0 KiB
  - taskethlw1c** 16,0 KiB
  - taskethlw1d 16,0 KiB
  - taskethlw1e 16,0 KiB
  - taskethlw1f 16,0 KiB
- sys
- world

spark.taskethlw1c: 5 строк (точно)

#	Employee_ID	Job_Code	City_code
1	E001	J01	26
2	E001	J02	26
3	E002	J02	56
4	E002	J03	56
5	E003	J01	56

200 SELECT \* FROM information\_schema.REFERENTIAL\_CONSTRAINTS WHERE CONSTRAINT\_SCHEMA='spark' AND TABLE\_NAME='taskethlw1c' AND REFERENCED\_TABLE\_NAME IS NOT NULL; ^

201 SELECT \* FROM information\_schema.KEY\_COLUMN\_USAGE WHERE TABLE\_SCHEMA='spark' AND TABLE\_NAME='taskethlw1c' AND REFERENCED\_TABLE\_NAME IS NOT NULL; ^

202 SHOW CREATE TABLE `spark`.`taskethlw1c`;

203 SELECT tc.CONSTRAINT\_NAME, cc.CHECK\_CLAUSE FROM `information\_schema`.`CHECK\_CONSTRAINTS` AS cc, `information\_schema`.`TABLE\_CONSTRAINTS` AS tc WHERE tc.CONSTRAINT

204 SELECT \* FROM `spark`.`taskethlw1c` LIMIT 1000;

r1: c2 Подключено: 23:0 MySQL 9.0.1 Время работы: 1 дней, 02:53 | Серверное время: Ожидание.

Unamed-1\spark\taskethw1d\ - HeidiSQL 12.8.0.6908

Файл Редактировать Поиск Запрос Инструменты Переход Помощь

Фильтр баз даннь Фильтр таблиц

Unamed-1 База данных: spark Таблица: taskethw1d Данные Запрос

Unamed-1

- information\_schema
- mysql
- performance\_schema
- sakila
- spark 144,0 KiB
  - tasketh1a 16,0 KiB
  - tasketh1b 16,0 KiB
  - tasketh1d 16,0 KiB
  - tasketh1c 16,0 KiB
  - taskethlw1b 16,0 KiB
  - taskethlw1c 16,0 KiB
  - taskethlw1d** 16,0 KiB
  - taskethlw1e 16,0 KiB
  - taskethlw1f 16,0 KiB
- sys
- world

spark.taskethw1d: 3 строк (точно)

#	Employee_ID	Name
1	E001	Alice
2	E002	Bob
3	E003	Alice

182 SHOW CREATE TABLE `spark`.`taskethlw1f`;

183 SELECT tc.CONSTRAINT\_NAME, cc.CHECK\_CLAUSE FROM `information\_schema`.`CHECK\_CONSTRAINTS` AS cc, `information\_schema`.`TABLE\_CONSTRAINTS` AS tc WHERE tc.CONSTRAINT

184 SELECT \* FROM `spark`.`taskethlw1f` LIMIT 1000;

185 SHOW CREATE TABLE `spark`.`taskethlw1d`;

186 SELECT \* FROM `spark`.`taskethlw1d` LIMIT 1000;

r1: c2 Подключено: 00:3 MySQL 9.0.1 Время работы: 1 дней, 05:50 | Серверное время: Ожидание.



Unamed-1\spark\tasketlhw1e - HeidiSQL 12.8.0.6908

Файл Редактировать Поиск Запрос Инструменты Переход Помощь

Фильтр баз данныФильтр таблиц

Unamed-1База данных: sparkТаблица: tasketlhw1eДанныеЗапрос

spark.tasketlhw1e: 3 строк (точно)ДалееПоказать всеСортировкаСтолбцы (2/2)Фильтр

#	Job_Code	Job
1	J01	Chef
2	J02	Waiter
3	J03	Bartender

ФильтРегулярное выражение

184 SELECT \* FROM `spark`.`tasketlhw1f` LIMIT 1000;

185 SHOW CREATE TABLE `spark`.`tasketlhw1d`;

186 SELECT \* FROM `spark`.`tasketlhw1d` LIMIT 1000;

187 SHOW CREATE TABLE `spark`.`tasketlhw1e`;

188 SELECT \* FROM `spark`.`tasketlhw1e` LIMIT 1000;

r1 : c2Подключено: 00:3MySQL 9.0.1Время работы: 1 дней, 05:53Серверное время: Ожидание.

Unamed-1\spark\tasketlhw1f - HeidiSQL 12.8.0.6908

Файл Редактировать Поиск Запрос Инструменты Переход Помощь

Фильтр баз данныФильтр таблиц

Unamed-1База данных: sparkТаблица: tasketlhw1fДанныеЗапрос

spark.tasketlhw1f: 2 строк (точно)ДалееПоказать всеСортировкаСтолбцы (2/2)Фильтр

#	City_code	Home_city
1	26	Moscow
2	56	Perm

ФильтРегулярное выражение

186 SELECT \* FROM `spark`.`tasketlhw1d` LIMIT 1000;

187 SHOW CREATE TABLE `spark`.`tasketlhw1e`;

188 SELECT \* FROM `spark`.`tasketlhw1e` LIMIT 1000;

189 SHOW CREATE TABLE `spark`.`tasketlhw1f`;

190 SELECT \* FROM `spark`.`tasketlhw1f` LIMIT 1000;

r1 : c2Подключено: 00:3MySQL 9.0.1Время работы: 1 дней, 05:53Серверное время: Ожидание.

