

cmd. Командная строка

Microsoft Windows [Version 10.0.19045.4412]

(c) Корпорация Майкрософт (Microsoft Corporation). Все права защищены.

C:\Users\Dom>chcp 65001 && spark-shell -i \spark\d1.scala --conf "spark.driver.extraJavaOptions=-Dfile.encoding=utf-8"

Active code page: 65001

Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties

Setting default log level to "WARN".

To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).

Spark context Web UI available at http://DESKTOP-TIN8UGG:4040

Spark context available as 'sc' (master = local[*], app id = local-1716024489547).

Spark session available as 'spark'.

ERROR StatusLogger Log4j2 could not find a logging implementation. Please add log4j-core to the classpath. Using SimpleLogger to log to the console...

```
+-----+-----+-----+-----+-----+
|Employee_ID| Name|Job_Code|      Job|City_code|Home_city|
+-----+-----+-----+-----+-----+
|      E001|Alice|   J01|   Chef|      26|  Moscow|
|      E001|Alice|   J02| Waiter|      26|  Moscow|
|      E002|  Bob|   J02| Waiter|      56|   Perm|
|      E002|  Bob|   J03|Bartender|      56|   Perm|
|      E003|Alice|   J01|   Chef|      56|   Perm|
+-----+-----+-----+-----+-----+
```

drop table spark.tabr1 complete

drop table spark.tabr2 complete

drop table spark.tabr3 complete

drop table spark.tabr4 complete

```
CREATE TABLE `tabr1` (
  `City_code` INT(10) NOT NULL,
  `Home_city` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
  PRIMARY KEY (`City_code`) USING BTREE
)
COMMENT='Успешно!'
COLLATE='utf8mb4_0900_ai_ci'
ENGINE=InnoDB
complete
```

```
CREATE TABLE `tabr2` (
  `Job_Code` VARCHAR(50) NOT NULL COLLATE 'utf8mb4_0900_ai_ci',
  `Job` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
  PRIMARY KEY (`Job_Code`) USING BTREE
)
COMMENT='Успешно!'
COLLATE='utf8mb4_0900_ai_ci'
ENGINE=InnoDB
;
complete
```

```
CREATE TABLE `tabr3` (
  `Employee_ID` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
  `Name` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
  `Job_Code` VARCHAR(50) NOT NULL COLLATE 'utf8mb4_0900_ai_ci',
  INDEX `Employee_ID` (`Employee_ID`, `Name`) USING BTREE,
  INDEX `Job_Code` (`Job_Code`) USING BTREE
)
COMMENT='Успешно!'
COLLATE='utf8mb4_0900_ai_ci'
ENGINE=InnoDB
;
complete
```

```
CREATE TABLE `tabr4` (
  `Employee_ID` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
  `Name` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
  `City_code` INT(10) NOT NULL,
  INDEX `Employee_ID` (`Employee_ID`, `Name`) USING BTREE,
  INDEX `City_code` (`City_code`) USING BTREE
)
```

```

        INDEX `City_code` (`City_code`) USING BTREE
    )
    COMMENT='Успешно!'
    COLLATE='utf8mb4_0900_ai_ci'
    ENGINE=InnoDB
;
complete

```

24/05/18 12:28:16 WARN ExcelHeaderChecker: Number of column in Excel header is not equal to number of fields in the schema:
Header length: 6, schema size: 2

Excel file: file:///spark/d1.xlsx

24/05/18 12:28:16 WARN ExcelHeaderChecker: Number of column in Excel header is not equal to number of fields in the schema:
Header length: 6, schema size: 2

Excel file: file:///spark/d1.xlsx

24/05/18 12:28:17 WARN ExcelHeaderChecker: Number of column in Excel header is not equal to number of fields in the schema:
Header length: 6, schema size: 3

Excel file: file:///spark/d1.xlsx

24/05/18 12:28:17 WARN ExcelHeaderChecker: Number of column in Excel header is not equal to number of fields in the schema:
Header length: 6, schema size: 3

Excel file: file:///spark/d1.xlsx

task 1

00:00:05

C:\Users\Dom>

	A	B	C	D	E	F	G
1	Employee_	Name	Job_Code	Job	City_code	Home_city	
2	E001	Alice	J01	Chef	26	Moscow	
3	E001	Alice	J02	Waiter	26	Moscow	
4	E002	Bob	J02	Waiter	56	Perm	
5	E002	Bob	J03	Bartender	56	Perm	
6	E003	Alice	J01	Chef	56	Perm	
7							

```

1  /*
2  |  chcp 65001 && spark-shell -i \spark\d1.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 tl = 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("/spark/d1.xlsx")
23         df1.show()
24
25         df1.filter(df1("Employee_ID").isNotNull).select("Employee_ID", "Name", "Job_Code", "Job", "City_code", "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", "tabr")
28         .mode("overwrite").save()
29
30         var q1 = ""
31         CREATE TABLE `tabr1` (
32             `City_code` INT(10) NOT NULL,
33             `Home_city` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
34             PRIMARY KEY (`City_code`) USING BTREE
35         )
36         COMMENT='Успешно!'
37         COLLATE='utf8mb4_0900_ai_ci'
38         ENGINE=InnoDB
39         ;
40         CREATE TABLE `tabr2` (
41             `Job_Code` VARCHAR(50) NOT NULL COLLATE 'utf8mb4_0900_ai_ci',
42             `Job` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
43             PRIMARY KEY (`Job_Code`) USING BTREE
44         )
45         COMMENT='Успешно!'
46         COLLATE='utf8mb4_0900_ai_ci'
47         ENGINE=InnoDB
48         ;
49         ""
50         import java.sql._;
51         def sqlexecute(sql: String) = {
52             var conn: Connection = null;
53             var stmt: Statement = null;
54             try {
55                 Class.forName("com.mysql.cj.jdbc.Driver");
56                 conn = DriverManager.getConnection("jdbc:mysql://localhost:3306/spark?user=root&password=root");
57                 stmt = conn.createStatement();
58                 stmt.executeUpdate(sql);
59                 println(sql+" complete");
60             } catch {
61                 case e: Exception => println ("exception caught" + e);
62             }

```



```

61         case e: Exception => println ("exception caught" + e);
62     }
63 }
64 sqlexecute("drop table spark.tabr1")
65 sqlexecute("drop table spark.tabr2")
66 sqlexecute("drop table spark.tabr3")
67 sqlexecute("drop table spark.tabr4")
68 sqlexecute("""
69 CREATE TABLE `tabr1` (
70     `City_code` INT(10) NOT NULL,
71     `Home_city` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
72     PRIMARY KEY (`City_code`) USING BTREE
73 )
74 COMMENT='Успешно!'
75 COLLATE='utf8mb4_0900_ai_ci'
76 ENGINE=InnoDB
77 """)
78 sqlexecute("""
79 CREATE TABLE `tabr2` (
80     `Job_Code` VARCHAR(50) NOT NULL COLLATE 'utf8mb4_0900_ai_ci',
81     `Job` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
82     PRIMARY KEY (`Job_Code`) USING BTREE
83 )
84 COMMENT='Успешно!'
85 COLLATE='utf8mb4_0900_ai_ci'
86 ENGINE=InnoDB
87 ;
88 """)
89 sqlexecute("""
90 CREATE TABLE `tabr3` (
91     `Employee_ID` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
92     `Name` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
93     `Job_Code` VARCHAR(50) NOT NULL COLLATE 'utf8mb4_0900_ai_ci',
94     INDEX `Employee_ID` (`Employee_ID`, `Name`) USING BTREE,
95     INDEX `Job_Code` (`Job_Code`) USING BTREE
96 )
97 COMMENT='Успешно!'
98 COLLATE='utf8mb4_0900_ai_ci'
99 ENGINE=InnoDB
100 ;
101 """)
102 sqlexecute("""
103 CREATE TABLE `tabr4` (
104     `Employee_ID` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
105     `Name` VARCHAR(50) NULL DEFAULT NULL COLLATE 'utf8mb4_0900_ai_ci',
106     `City_code` INT(10) NOT NULL,
107     INDEX `Employee_ID` (`Employee_ID`, `Name`) USING BTREE,
108     INDEX `City_code` (`City_code`) USING BTREE
109 )
110 COMMENT='Успешно!'
111 COLLATE='utf8mb4_0900_ai_ci'
112 ENGINE=InnoDB
113 ;
114 """)
115 df1.select("City_code","Home_city").distinct()
116 .write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
117 .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tabr1")
118 .mode("append").save()
119 df1.select("Job_Code","Job").distinct()
120 .write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
121 .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tabr2")
122 .mode("append").save()

```

```
121     .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tabr2")
122     .mode("append").save()
123     df1.select("Employee_ID","Name","Job_Code").distinct()
124     .write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
125     .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tabr3")
126     .mode("append").save()
127     df1.select("Employee_ID","Name","City_code").distinct()
128     .write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
129     .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "tabr4")
130     .mode("append").save()
131
132     println("task 1")
133 -}
134 val s0 = (System.currentTimeMillis() - t1)/1000
135 val s = s0 % 60
136 val m = (s0/60) % 60
137 val h = (s0/60/60) % 24
138 println("%02d:%02d:%02d".format(h, m, s))
139 System.exit(0)
```



Фильтр баз д... Фильтр табли...

Unnamed База данных: spark Таблица: tabr Данные Запрос

▼ Unnamed

- > information_sc...
- > mysql
- > performance_s...

▼ spark 144,0 KiB

- tabr 16,0 KiB
- tabr1 16,0 KiB
- tabr2 16,0 KiB
- tabr3 48,0 KiB
- tabr4 48,0 KiB

> sys

spark.tabr: 5 строк (точно)

#	Employee_ID	Name	Job_Code	Job	City_code	Home_city
1	E001	Alice	J01	Chef	26	Moscow
2	E001	Alice	J02	Waiter	26	Moscow
3	E002	Bob	J02	Waiter	56	Perm
4	E002	Bob	J03	Bartender	56	Perm
5	E003	Alice	J01	Chef	56	Perm



Фильтр баз д... Фильтр табли...



Unnamed

База данных: spark



▼ Unnamed

> information_sc...

> mysql

> performance_s...

▼ spark 144,0 KiB

tabr 16,0 KiB

tabr1 16,0 KiB

tabr2 16,0 KiB

tabr3 48,0 KiB

tabr4 48,0 KiB

> sys

spark.tabr1: 2 строк (точно)

#	City_code	Home_city
1	26	Moscow
2	56	Perm



▼ **Unnamed**

- > information_sc...
- > mysql
- > performance_s...
- ▼ **spark** 144,0 KiB
 - tabr 16,0 KiB
 - tabr1 16,0 KiB
 - tabr2** 16,0 KiB
 - tabr3 48,0 KiB
 - tabr4 48,0 KiB
- > sys

spark.tabr2: 3 строк (точно)

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



▼ Unnamed

- > information_sc...
- > mysql
- > performance_s...
- ▼ **spark** 144,0 KiB
 - tabr 16,0 KiB
 - tabr1 16,0 KiB
 - tabr2 16,0 KiB
 - ★ **tabr3** 48,0 KiB
 - tabr4 48,0 KiB
- > sys

spark.tabr3: 5 строк (точно)

#	Employee_ID	Name	Job_Code
1	E001	Alice	J01
2	E001	Alice	J02
3	E002	Bob	J02
4	E002	Bob	J03
5	E003	Alice	J01



▼ Unnamed

- > information_sc...
- > mysql
- > performance_s...
- ▼ spark 144,0 KiB
 - tabr 16,0 KiB
 - tabr1 16,0 KiB
 - tabr2 16,0 KiB
 - tabr3 48,0 KiB
 - tabr4 48,0 KiB**
- > sys

spark.tabr4: 3 строк (точно)

#	Employee_ID	Name	City_code
1	E001	Alice	26
2	E002	Bob	56
3	E003	Alice	56