Spark直接访问MySQL数据库

spark可以直接访问MySQL业务数据库，获取数据，但是这样可能会导致业务数据库压力比较大，数据量较大时，不推荐这样处理

In [1]:
```python
import os
# 配置pyspark和spark driver运行时 使用的python解释器
JAVA_HOME = '/root/bigdata/jdk'
PYSPARK_PYTHON = '/miniconda2/envs/py365/bin/python'
# 当存在多个版本时，不指定很可能会导致出错
os.environ['PYSPARK_PYTHON'] = PYSPARK_PYTHON
os.environ['PYSPARK_DRIVER_PYTHON'] = PYSPARK_PYTHON
os.environ['JAVA_HOME'] = JAVA_HOME
# 实现把mysql-connector-java-5.1.47.jar放入指定目录下
# 如果是使用spark直连MySQL数据库，读取数据，需要使用mysql链接的jar包，这里预先下载并放
os.environ["PYSPARK_SUBMIT_ARGS"] = "--jars /root/bigdata/spark/extraJars/mysql-conne
# 配置spark信息
from pyspark import SparkConf
from pyspark.sql import SparkSession

SPARK_APP_NAME = "preprocessingMeiDuoData"
SPARK_URL = "spark://192.168.58.100:7077"

conf = SparkConf()      # 创建spark config对象
config = (
    ("spark.app.name", SPARK_APP_NAME),      # 设置启动的spark的app名称，没有提供，将随村
    ("spark.executor.memory", "2g"),     # 设置该app启动时占用的内存用量，默认1g，指一:
    ("spark.master", SPARK_URL),      # spark master的地址
    ("spark.executor.cores", "2"),      # 设置spark executor使用的CPU核心数，指一台虚拟村
    # 以下三项配置，可以控制执行器数量
#     ("spark.dynamicAllocation.enabled", True),
#     ("spark.dynamicAllocation.initialExecutors", 1),     # 1个执行器
#     ("spark.shuffle.service.enabled", True)
#   ('spark.sql.pivotMaxValues', '99999'),  # 当需要pivot DF，且值很多时，需要修改，默
)
# 查看更详细配置及说明：https://spark.apache.org/docs/latest/configuration.html

conf.setAll(config)

# 利用config对象，创建spark session
spark = SparkSession.builder.config(conf=conf).enableHiveSupport().getOrCreate()
```

```
In [4]:  1  # 直连访问mysql读取数据，这种方式数据没有经过hdfs，而是直接在MySQL数据库上进行操作，因
         2  # 因此该方式虽然用起来最爽，但是使用时一定要注意对mysql数据带来的影响
         3  ret = spark.read.format("jdbc").options(url="jdbc:mysql://localhost/meiduo_mall",
         4                                          driver="com.mysql.jdbc.Driver",
         5                                          dbtable="tb_goods",
         6                                           user="root",
         7                                          password="password").load()#mysql的密码
         8  ret
```

Out[4]: DataFrame[id: int, create_time: timestamp, update_time: timestamp, name: string, sales:
         int, comments: int, brand_id: int, category1_id: int, category2_id: int, category3_id:
         int, desc_detail: string, desc_pack: string, desc_service: string]

In [5]:
```
1 ret.show()
```

```
+---+--------------------+--------------------+--------------------+-----+--------+--------------------+--------------------+
| id|         create_time|         update_time|                name|sales|comments|brand_id|category1_id|category2_id|category3_id|         desc_detail|         desc_pack|         desc_service|
+---+--------------------+--------------------+--------------------+-----+--------+--------------------+--------------------+
|  1|2018-04-11 16:01:...|2018-04-25 12:09:...|Apple MacBook Pro...|    1|       1|       1|           4|          45|         157|<h1 style="text-a...|<h2>包装清单</h2>
...|<p> <strong>...|
|  2|2018-04-14 02:09:...|2018-04-25 11:51:...| Apple iPhone 8 Plus|    3|       1|       1|           1|          38|         115|<p><img alt="" sr...|<h3>包装清单</h3>
...|<p> <strong>...|
|  3|2018-04-14 03:03:...|2018-04-25 11:51:...|  华为 HUAWEI P10 Plus|    1|       8|       2|           1|          38|         115|<p><img alt="" sr...|<h3>包装清单</h3>
...|<p> <strong>...|
|  4| 2018-11-01 10:43:14| 2018-11-01 10:43:14|【次日达】泰火薄快充电宝手机壳无线...|    0|       0|       1|           1|          39|         126|                    |                    |
|  5| 2018-11-01 10:43:25| 2018-11-01 10:43:25|GoPro hero7运动相机水下...|    0|       0|       1|           2|          40|         132|                    |                    |
|  6| 2018-11-01 10:43:58| 2018-11-01 10:43:58|川宇 USB3.0高速多功能合一T...|    0|       0|       1|           3|          41|         140|                    |                    |
|  7| 2018-11-01 10:43:59| 2018-11-01 10:43:59|腾讯听听 9420　智能音箱/音...|    0|       0|       1|           3|          42|         143|                    |                    |
|  8| 2018-11-01 10:44:00| 2018-11-01 10:44:00|苹果XS Max无线充电器iPho...|    0|       0|       1|           1|          39|         123|                    |                    |
|  9| 2018-11-01 10:44:05| 2018-11-01 10:44:05|behringer 百灵达X32数...|    0|       0|       1|           3|          42|         146|                    |                    |
| 10| 2018-11-01 10:44:13| 2018-11-01 10:44:13|ESCASE 荣耀畅玩8c手机壳 ...|    0|       0|       1|           1|          39|         119|                    |                    |
| 11| 2018-11-01 10:44:11| 2018-11-01 10:44:11|FDY FW3C 5C 蓝牙无线W...|    0|       0|       1|           3|          42|         143|                    |                    |
| 12| 2018-11-01 10:44:12| 2018-11-01 10:44:12|摩托罗拉（Motorola） 威泰...|    0|       0|       1|           1|          38|         118|                    |                    |
| 13| 2018-11-01 10:44:14| 2018-11-01 10:44:14|沃丁（WODING） 亚马逊 Am...|    0|       0|       1|           3|          42|         143|                    |                    |
| 14| 2018-11-01 10:44:16| 2018-11-01 10:44:16|科凌（keling） F8全波段收...|    0|       0|       1|           3|          42|         144|                    |                    |
| 15| 2018-11-01 10:44:25| 2018-11-01 10:44:25|科威盛（kevsen） Q3对讲机...|    0|
```

```
   0|          1|          1|        38|       118|                    |
   |                    |
| 16| 2018-11-01 10:44:24| 2018-11-01 10:44:24|小米（MI） 小米手环3代nfc版...|     0|
   0|          1|          3|        43|       147|                    |
   |                    |
| 17| 2018-11-01 10:44:25| 2018-11-01 10:44:25|摩托罗拉（Motorola） MA...|     0|
   0|          1|          1|        38|       118|                    |
   |                    |
| 18| 2018-11-01 10:44:29| 2018-11-01 10:44:29|C&C 儿童相机数码卡通照相机玩具...|
   0|          0|          1|         2|        40|       128|                    |
   |                    |
| 19| 2018-11-01 10:44:30| 2018-11-01 10:44:30|全景看房婚礼旅行！理光 THETA...|     0|
   0|          1|          2|        40|       128|                    |
   |                    |
| 20| 2018-11-01 10:44:29| 2018-11-01 10:44:29|欧兴 KLG5.0 对讲机 待机1...|     0|
   0|          1|          1|        38|       118|                    |
   |                    |
+---+-------------------+-------------------+--------------------+-----+-------+--
-------+------------+------------+-----------+----------+--------------------+----
---+--------------------+

only showing top 20 rows
```