实验3

上传数据：

```
hadoop fs -put /usr/local/src/jizhan_information.csv /user/root/data/jz.csv
```

在 hive 的 card 数据库中创建表格：

```
CREATE TABLE jz (
    record_time string,
    imei int,
    cell string,
    ph_num int,
    call_num int,
    drop_num int,
    duration int,
    drop_rate double,
    net_type string,
    erl int)
    row format delimited fields terminated by ','
;
```

装载数据：

```
# 本地文件TODO:
load data local inpath '' overwrite into table jz;
# HDFS中文件
load data inpath '/user/local/src/jizhan_information.csv' overwrite into
table jz;
```

执行命令：

```
SELECT imei, SUM(drop_num)/SUM(duration)*100 AS droprate
FROM jz
GROUP BY imei
ORDER BY droprate
desc limit 10;
```

发现报错：

错误: 找不到或无法加载主类org.apache.hadoop.mapreduce.v2.app.MRAppMaster

解决方法：

在命令行下输入如下命令，并将返回的地址复制：

```
hadoop classpath
```

编辑 `yarn-site.xml`

```
vim /usr/local/hadoop/etc/hadoop/yarn-site.xml
```

添加下面内容

```
<configuration>
    <property>
        <name>yarn.application.classpath</name>
        <value>输入刚才返回的Hadoop classpath路径</value>
    </property>
</configuration>
```

我的是下面这样

```
    <property>
        <name>yarn.application.classpath</name>

<value>/usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/common/*:/usr/local/hadoop/share/hadoop/hdfs:/usr/local/hadoop/share/hadoop/hdfs/lib/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/local/hadoop/share/hadoop/mapreduce/*:/usr/local/hadoop/share/hadoop/yarn:/usr/local/hadoop/share/hadoop/yarn/lib/*:/usr/local/hadoop/share/hadoop/yarn/*
</value>
    </property>
```

再次执行命令：

```
hive> SELECT imei, SUM(drop_num)/SUM(duration)*100 AS droprate
    > FROM jz
    > GROUP BY imei
    > ORDER BY droprate
    > desc limit 10;
Query ID = root_20230520094450_6abaef41-b50b-4977-b6d3-ebb817c645f8
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1684543101265_0005, Tracking URL = http://master:8088/proxy/application_1684543101265_0005/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1684543101265_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-20 09:45:17,265 Stage-1 map = 0%,  reduce = 0%
2023-05-20 09:45:29,971 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.43 sec
2023-05-20 09:45:42,991 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 10.5 sec
MapReduce Total cumulative CPU time: 10 seconds 500 msec
Ended Job = job_1684543101265_0005
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1684543101265_0006, Tracking URL = http://master:8088/proxy/application_1684543101265_0006/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1684543101265_0006
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-05-20 09:46:02,942 Stage-2 map = 0%,  reduce = 0%
2023-05-20 09:46:14,009 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 3.2 sec
2023-05-20 09:46:23,639 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 6.91 sec
MapReduce Total cumulative CPU time: 6 seconds 910 msec
Ended Job = job_1684543101265_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 10.5 sec   HDFS Read: 57416065 HDFS Write: 334905 SUCCESS
```

最终结果

```
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1684334891813_0007, Tracking URL = http://master:8088/proxy/application_1684334891813_0007/
Kill Command = /usr/local/hadoop/bin/mapred job  -kill job_1684334891813_0007
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-05-17 23:07:08,915 Stage-2 map = 0%,  reduce = 0%
2023-05-17 23:07:16,569 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 3.27 sec
2023-05-17 23:07:23,796 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 6.03 sec
MapReduce Total cumulative CPU time: 6 seconds 30 msec
Ended Job = job_1684334891813_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.29 sec   HDFS Read: 57415927 HDFS Write: 334905 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 6.03 sec   HDFS Read: 342334 HDFS Write: 487 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 320 msec
OK
639876  0.0013623978201634877
356436  9.727626459143969E-4
351760  8.116883116883117E-4
368883  6.906077348066298E-4
358849  6.807351940095302E-4
358231  6.199628022318661E-4
863738  5.982650314089142E-4
865011  5.36480686695279E-4
862242  5.227391531625719E-4
350301  5.002501250625312E-4
Time taken: 60.69 seconds, Fetched: 10 row(s)
```