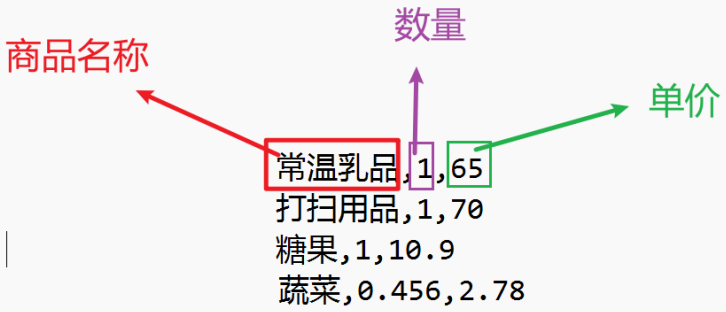
7.1.7 Hive映射测试

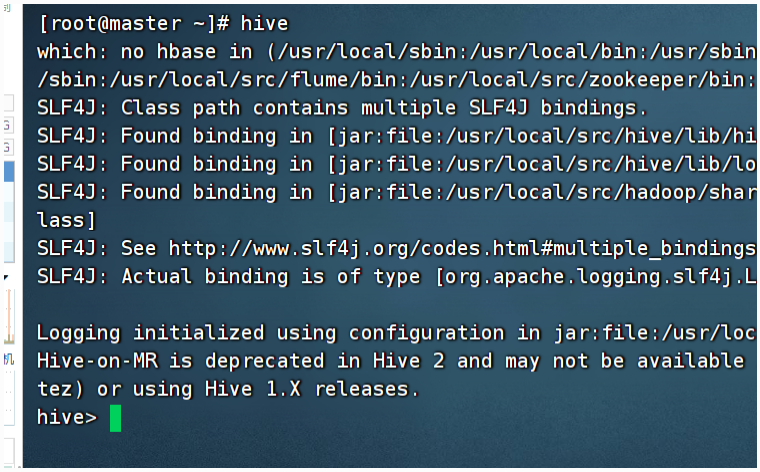
以MR章节的销售额数据为例

我们以之前在MapReduce对sale.csv做的营业额分析为例，我们可以通过Hive的HQL语句快速的完成MR计算。



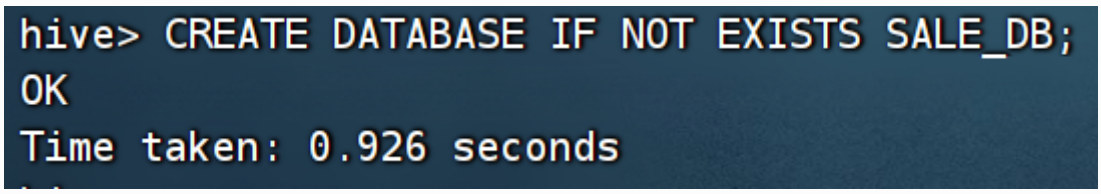
打开hive

```
hive
```



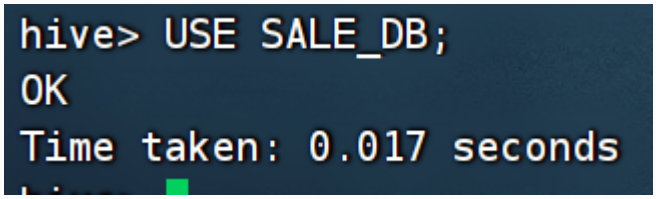
创建SALE_DB数据库

```
hive> CREATE DATABASE IF NOT EXISTS SALE_DB;
```



切换SALE_DB数据库

```
hive> USE SALE_DB;
```



创建SALE表

```
create table sale (
    goodType varchar(50),
    volume float,
    unit float
) row format delimited fields terminated by ","
lines terminated by "\n"
stored as textfile;
```

```
hive> create table sale (
    > goodType string
    > ,volume float
    > ,unit float
    > )row format delimited fields terminated by ","
    > lines terminated by "\n"
    > stored as textfile;
OK
Time taken: 0.418 seconds
```

此时我们来到 https://master:50070，看到Hive创建了如下层级的目录(/user/hive/warehouse/sale_db.db/sale/

Browse Directory

/user/hive/warehouse/sale_db.db/sale Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name

我们先将sale.csv文件上传到hdfs上的/user/hive/warehouse/sale_db.db/sale/上

（**步骤省略，已经学到这了，该自己写了**）上传结果如下：

Browse Directory

/user/hive/warehouse/sale_db.db/sale Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	root	supergroup	787.56 KB	2023/11/20 21:11:13	1	128 MB	sale.csv

Hadoop, 2015.

在Hive中查看该数据

```
hive> select * from sale limit 5;
```

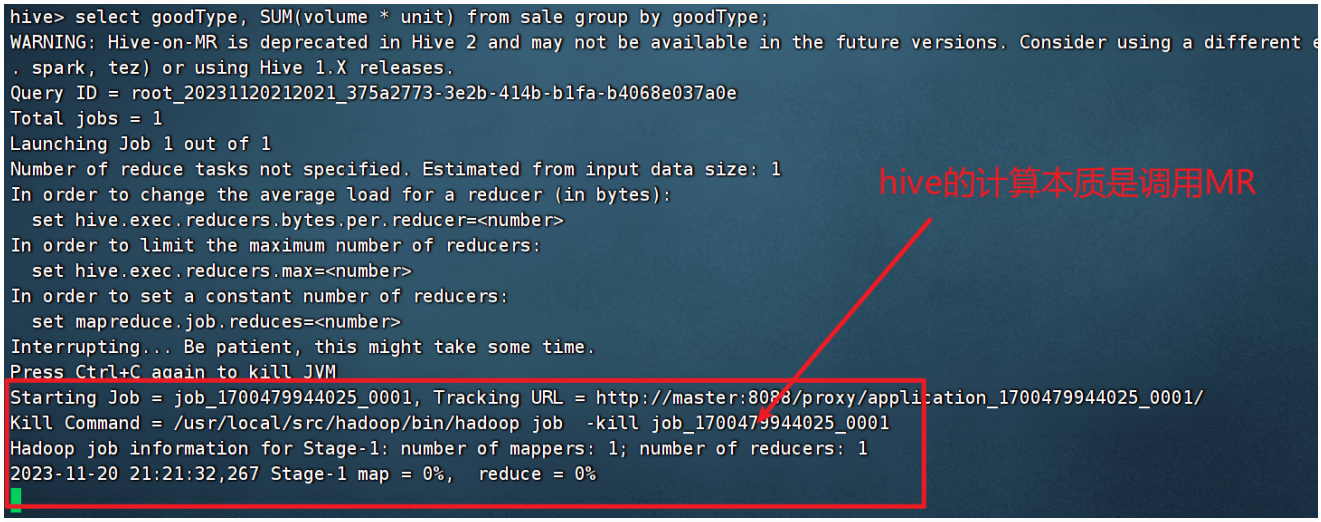
我们就可以发现该文本文件被**映射**成了表结构

```
hive> select * from sale limit 5;
OK
袋装速食面      1.0     3.0
调味酱  1.0     6.7
冷藏乳品        1.0     8.9
冲调食品        1.0     27.7
洗衣用品        1.0     4.9
Time taken: 0.101 seconds, Fetched: 5 row(s)
```

我们可以通过HQL语句完成对营业额的统计

```
hive> select goodType, SUM(volume * unit) from sale group by goodType;
```

启动MR进行计算



计算结果如下：

```
        NULL
一次性用品      822.0999979972839
不锈钢餐具      37.0
个人卫生用品    47.20000076293945
中式熟菜        477.40000581741333
乳饮料  2760.7000076770782
五谷杂粮        13339.44566068612
保养用品        1848.6999962329865
保温容器        181.0
保鲜用品        460.89999771118164
其他国产酒      585.9999990463257
其它加工        450.28600327670574
冰品    73.5
冰鲜水产        1528.036190778017
冲调食品        6427.499966144562
冲饮品  700.8999950885773
冷冻包子馒头    478.50000047683716
```

比如去设计MR的程序，使用我们更加熟悉的类似于SQL的**HQL语句**来实现对目标的统计计算，会更加简便，一句HQL语句就等于我们之前写的三个程序，这也是Hive的强大之处。

在接下来的章节，我们就要逐步来学习Hive和HQL语句的使用。