

Spark 综合案例



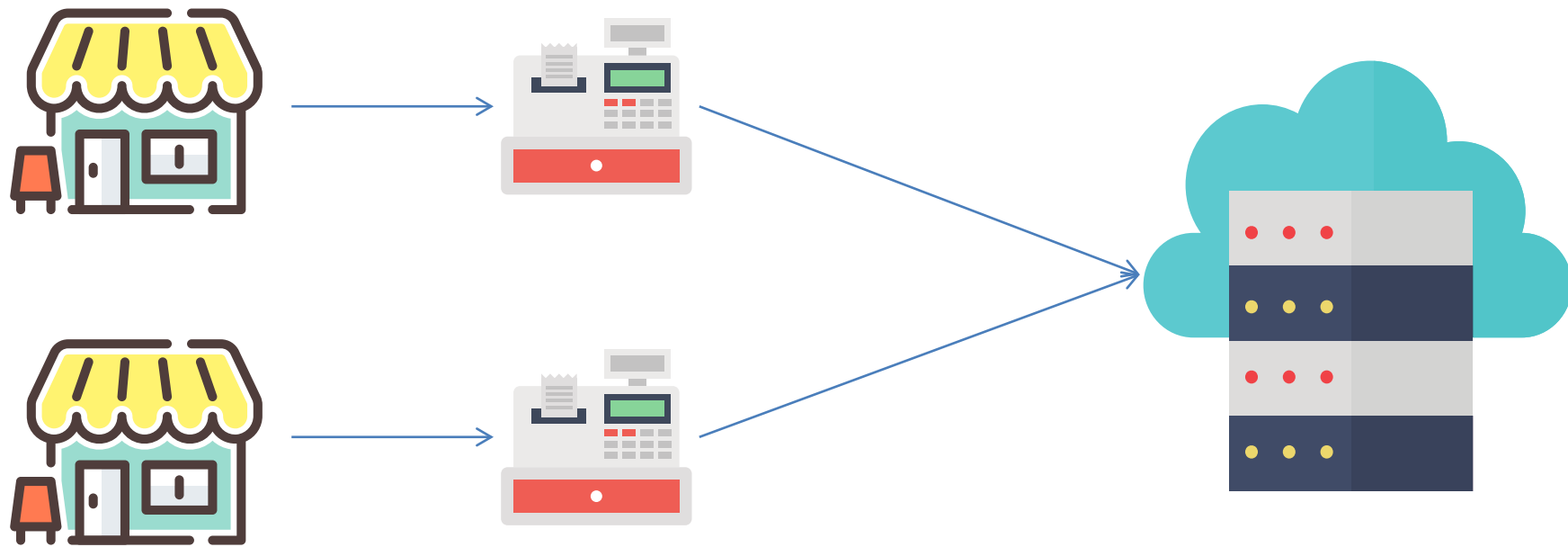
需求分析

某公司是做零售相关业务, 旗下出品各类收银机.

目前公司的收银机已经在全国铺开, 在各个省份均有店铺使用.

机器是联网的, 每一次使用都会将售卖商品数据上传到公司后台.

老板现在想对省份维度的销售情况进行统计分析



需求

4个需求需开发

1. 各省 销售 指标 每个省份的销售额统计
2. TOP3 销售省份中, 有多少家店铺 日均销售额 1000+
3. TOP3 省份中 各个省份的平均单单价
4. TOP3 省份中, 各个省份的支付类型比例

2个操作需处理

1. 将需求结果写出到mysql
2. 将数据写入到Spark On Hive中

4个需求中2个较简单,2个较复杂
比较合适对SparkSQL的内容进行综合训练

数据格式

```
JSON
{
  "discountRate": 1,
  "dayOrderSeq": 8,
  "storeDistrict": "雨花区",
  "isSigned": 0,
  "storeProvince": "湖南省",
  "origin": 0,
  "storeGPSLongitude": "113.01567856440359",
  "discount": 0,
  "storeID": 4064,
  "productCount": 4,
  "operatorName": "OperatorName",
  "operator": "NameStr",
  "storeStatus": "open",
  "storeOwnUserTel": 12345678910,
  "corporator": "hny",
  "serverSaved": true,
  "payType": "alipay",
  "discountType": 2,
  "storeName": "杨光峰南食店",
  "storeOwnUserName": "OwnUserNameStr",
  "dateTS": 1563758583000,
  "smallChange": 0,
  "storeGPSName": "",
  "erase": 0,
  "product": [
    {
      "count": 1,
      "name": "百事可乐可乐型汽水",
      "unitID": 0,
      "barcode": "6940159410029",
      "pricePer": 3,
      "retailPrice": 3,
      "tradePrice": 0,
      "categoryID": 1
    },
    {
      "count": 1,
      "name": "百事可乐可乐型汽水",
      "unitID": 0,
      "barcode": "6940159410029",
      "pricePer": 3,
      "retailPrice": 3,
      "tradePrice": 0,
      "categoryID": 1
    },
    {
      "count": 1,
      "name": "百事可乐可乐型汽水",
      "unitID": 0,
      "barcode": "6940159410029",
      "pricePer": 3,
      "retailPrice": 3,
      "tradePrice": 0,
      "categoryID": 1
    },
    {
      "count": 1,
      "name": "百事可乐可乐型汽水",
      "unitID": 0,
      "barcode": "6940159410029",
      "pricePer": 3,
      "retailPrice": 3,
      "tradePrice": 0,
      "categoryID": 1
    }
  ],
  "storeGPSAddress": "",
  "orderID": "156375858240940641230",
  "moneyBeforeWholeDiscount": 22.5,
  "storeCategory": "normal",
  "receivable": 22.5,
  "faceID": "",
  "storeOwnUserId": 4082,
  "paymentChannel": 0,
  "paymentScenarios": "PASV",
  "storeAddress": "StoreAddress",
  "totalNoDiscount": 22.5,
  "payedTotal": 22.5,
  "storeGPSLatitude": "28.121213726311993",
  "storeCreateDateTS": 1557733046000,
  "payStatus": -1,
  "storeCity": "长沙市",
  "memberID": "0"
}
```

常用字段解析:

- storeDistrict 店铺所在行政区
- **storeProvince** 店铺所在省份
- **storeID** 店铺ID
- storeName 店铺名称
- **dateTS** 订单日期
- orderID 订单ID
- **receivable** 收款金额
- **payType** 付款类型

尽管字段较多, 我们只需要用到上述5个红色字段即可

数据在课程资料中有所提供, 文件名: mini.json

现在进入开发吧