**MRS作业二操作指导书**

1. **作业二**

1.1 项目介绍

本次实践以虚拟案例来引导用户熟悉华为云MRS服务的使用。

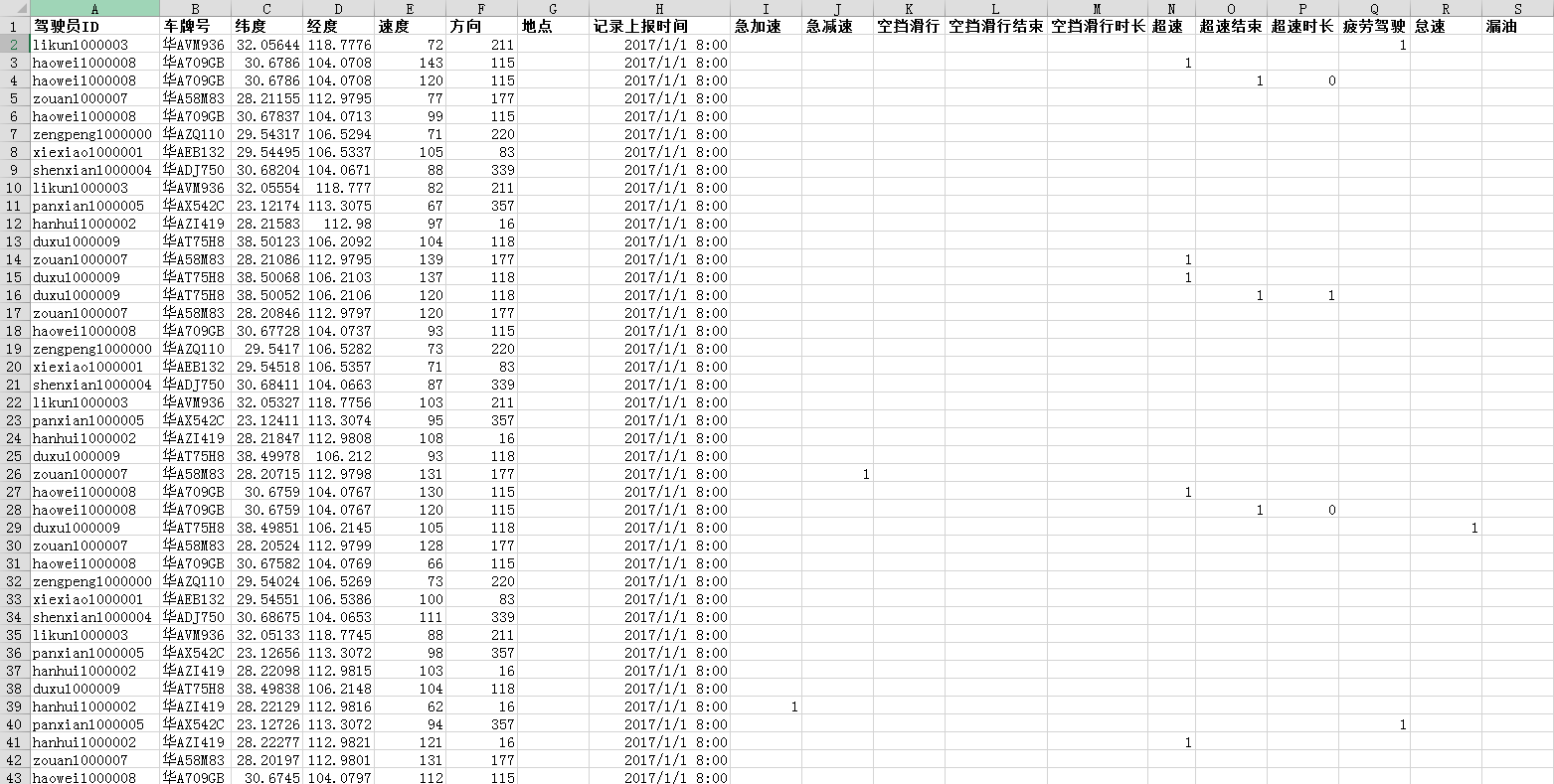
我们拥有一份用户驾驶过程中产生的信息数据，包括车主在日常的驾驶行为中是否急加速、急减速、空挡滑行、超速、疲劳驾驶等信息，然后通过MRS服务Spark组件分析统计指定时间内，车主急加速、急刹车、空挡滑行、超速、疲劳驾驶等违法行为的次数。

数据下载地址：<http://mapreduceservice.obs-website.cn-north-1.myhwclouds.com/demo/driver-analysis/detail-records.zip>

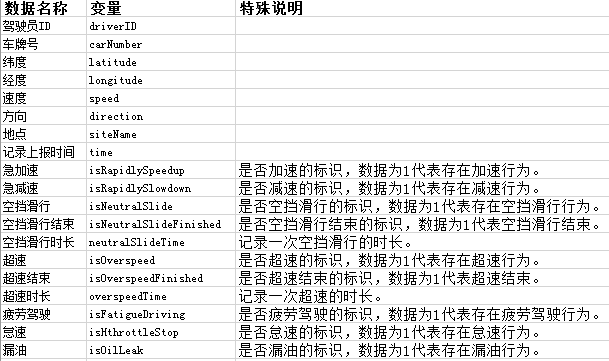
程序下载地址：<https://public-mrs.obs-website.cn-north-1.myhwclouds.com/driver_behavior.jar>

1.2 数据介绍

其中数据含义如下图，其中每行数据代表一次记录，违章行为的值如果为1就代表违章，如果为空就代表本次记录没有该违章行为。



数据定义和详解如下（作业三的变量名可参考下图）：

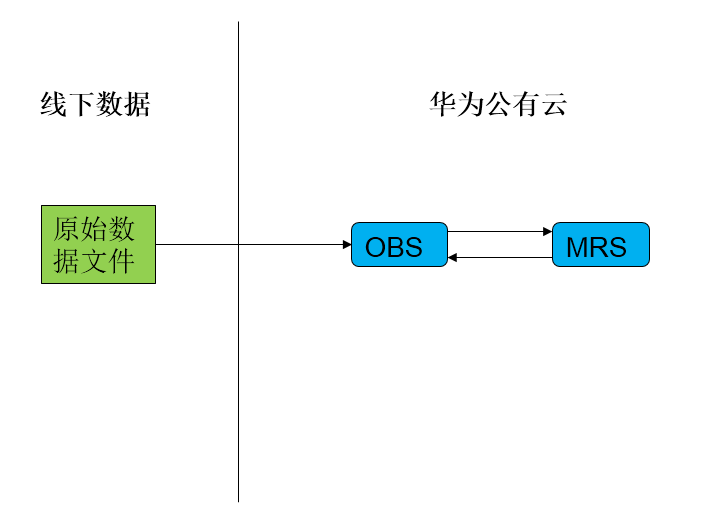


1.3 解决方案介绍

1.3.1 原始数据文件通过华为云的对象存储服务（OBS）从线下上传到云上。

1.3.2 提交Spark任务到MRS集群读取OBS中的数据，分析完成后将结果文件写回OBS。

1.3.3 从OBS上下载结果数据查看。



1.4 完成目标

最终通过分析完成的目标就是得到下列数据。



**2. 作业执行**

2.1 根据作业一操作指导书创建MRS集群

2.2 下载数据

2.2.1 从<https://public-mrs.obs-website.cn-north-1.myhwclouds.com/driver_behavior.jar>

路径下载样例程序driver\_behavior.jar至本地。

2.2.2 从<http://mapreduceservice.obs-website.cn-north-1.myhwclouds.com/demo/driver-analysis/detail-records.zip>

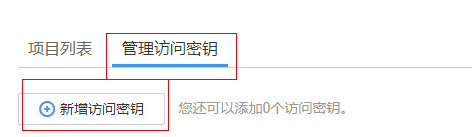
获取“detail-records.zip”压缩包，并进行解压。

2.3 创建对象存储

2.3.1 登录华为云控制台，在页面右上角的用户名下拉列表中点击“我的凭证”。



2.3.2 在“我的凭证”页面切换到“管理访问密钥”页签，单击“新增访问密钥”，在弹出的对话框中输入密码，然后点击“确定”。新增成功后会下载一个excel文件，打开后获取“Access Key Id”和“Secret Access Key”。



2.3.3 登录华为云控制台，选择“存储 >[对象存储服务](https://storage.huaweicloud.com/obs)”。在对象存储服务页面右上角点击“创建桶”。将桶名称修改为“obs-学校简称-姓名拼音”后点击下方的“立即创建”按钮。（因为桶名全局唯一，如果与其他人重复，桶名最后递增数字，例如obs-abcd-zhangsan1）



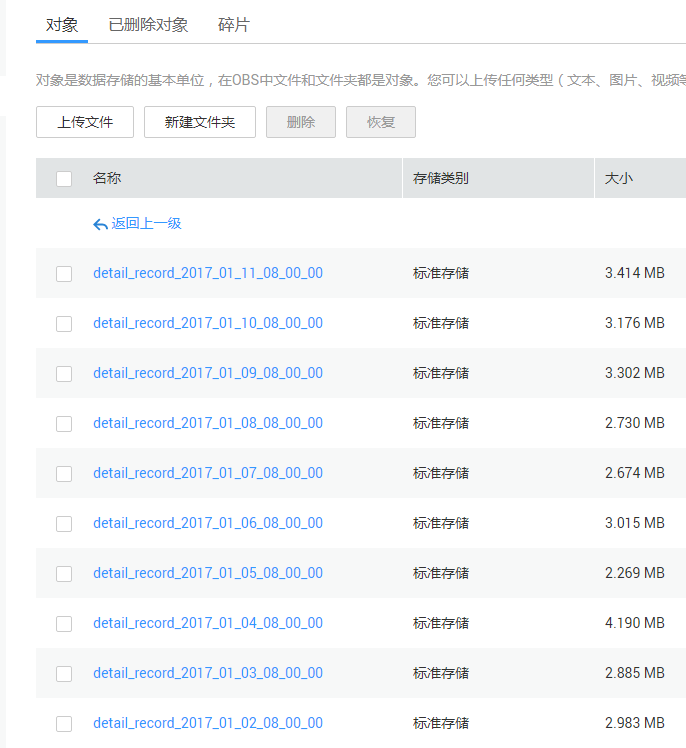
2.3.4 点击上述步骤中创建的桶名，然后在左树中选择“对象”，然后点击“新建文件夹”，输入文件夹的名字以创建文件夹。需要创建"input"、“output”两个文件夹。创建完成以后如下图所示：



2.3.5 然后点击“上传文件”，在弹出的对话框中选择下载的“driver\_behavior.jar”进行上传。



2.3.6点击创建的“input”文件夹，然后点击“上传文件”，在弹出的对话框中选择解压后的数据文件进行上传（由于当前OBS页面上传只支持单文件上传，故十个数据文件需要一一上传）。上传后结果如下图：

​

2.4 提交任务

2.4.1 点击[MRS管理界面](https://console.huaweicloud.com/console-emr/?region=cn-north-1#/manage/clusterList/clusterRunList)进入到集群界面。

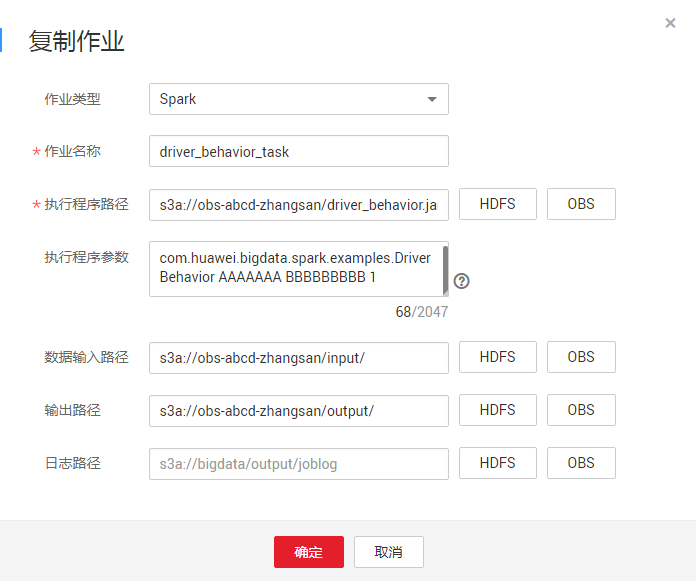
2.4.2 在列表中找到之前创建的MRS集群，等到集群状态变成“运行中”后，就可单击该集群名称，切换到“作业管理”页签，点击“添加”。



2.4.3 在弹出的对话框中，按照以下信息进行作业配置：

|  |  |
| --- | --- |
| 参数 | 值 |
| 作业类型 | Spark |
| 作业名称 | driver\_behavior\_task |
| 执行程序路径 | s3a://obs-abcd-zhangsan/driver\_behavior.jar |
| 执行程序参数 | com.huawei.bigdata.spark.examples.DriverBehavior AK SK 1  **其中AK和SK需要替换成之前获取到的Access Key Id和Secret Access Key。** |
| 数据输入路径 | s3a://obs-abcd-zhangsan/input/ |
| 输出路径 | s3a://obs-abcd-zhangsan/output/ |
| 日志路径 |  |

填写完成以后如下图所示：



2.4.4 单击“确定”按钮开始执行程序。稍等1-2分钟就可以到OBS中查看结果。

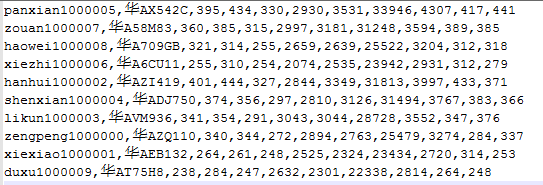
2.5 查看结果

2.5.1 登录[华为云控制台](https://console.huaweicloud.com/console/?region=cn-north-1#/home)，选择“存储 > [对象存储服务](https://storage.huaweicloud.com/obs/#/obs/buckets)”。

2.5.2 点击之前创建的桶名称“obs-abcd-zhangsan”，可以看到output文件夹中生成了一个以“作业名称”即“driver\_behavior\_task”结尾的文件夹，点击该文件夹即可看到执行结果，点击“下载”按钮将该文件下载到本地。



2.5.3 最终结果如下图，如果使用Excel表格打开，第二列会出现乱码，因为excel默认不支持UTF-8格式的中文显示，可使用导入的方式查看数据或者不做处理。



**3. 作业截图**

需要截出2.5.2步骤的图，并且显示出obs桶名，以确认具体完成人。

