**2020年安徽省大数据与人工智能应用竞赛A卷**

**学校名称： 队伍名称： 队伍编号：**

# 第六部分：综合题目（15分）

### 1、Mapreduce和Spark的都是并行计算，那么他们有什么相同和区别（5分）

答：

### 2、利用spark完成下面的统计分析（10分）

数据为某网站的访问日志：



字段分别为 ip 时间 操作方式以及访问地址 返回状态码 数据量字节数

源码：

1：计算独立IP数（2分）

结果截图：

代码截图：

2：统计每个图片独立IP数 （图片为地址中含有.png）（4分）

结果截图：

代码截图：

3：统计一天中每个小时的流量（4分）

结果截图：

代码截图：