# 预测项目处理流程

## 数据准备

## 从hdfs 获取源数据数据：

路径为:/new7g/result/minute/sta\_mi\_umac\_nber

具体路径为：/new7g/result/minute/sta\_mi\_umac\_nber/年/月/日/part-xxxxx

数据格式为：日期(YYYYMMDDHHmm),区块,分钟人流数，小时人流数，天人流数

## 二、数据清洗

1. **缺失数据补充 线性插值**
2. **异常值识别与剔除**

## 三、数据输出

**数据输出存储到Hbase 表：7g\_forecast\_umac\_nber**

小时预测数据格式：

Rowkey:时间\_区域块(YYYYMMDDHH\_Stype)

Value:json 格式化为 "mSum":value

分钟预测数据格式：

Rowkey:时间\_区域块(YYYYMMDDHH\_Stype\_MIN)

备注：MIN为固定标识符用于与按小时预测区分开

Value:json 格式化为 "mSum":value

## 四、算法

1、时间序列

2、指数平滑加权历史数据