

## 《快学 Spark 2.0》

# 第 33 讲、使用 updateStateByKey 完善案例

讲师-Cloudy

### 1.1 updateStateByKey

updateStateByKey :

以 DStream 中的数据进行按 key 做 reduce 操作，然后对各个批次的数据进行累加，输出总结果。

使用这个功能需要完成两步：

- 1) 定义状态：可以是任意数据类型。
- 2) 定义状态更新函数：用一个函数指定如何使用先前的状态，从输入流中的新值更新状态。

对于有状态操作，要不断的把当前和历史的时间切片的 RDD 累加计算。

当前批次的数据 Seq;

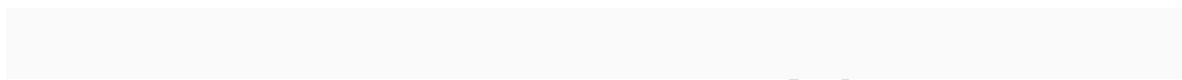
截至上个批次的结果 State;

汇总，就是总结果。

时间窗口、时间段（开始时间固定）

按自然日统计，wordCount，如何做？

按日期-地区统计销售额



www.ibEIFENG.com