



链滴

# 解决 spark streaming 集成 kafka 时只能读取每个 topic 的其中一个分区数据的问题

作者: [leekeggs](#)

原文链接: <https://ld246.com/article/1550565415640>

来源网站: 链滴

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

# 1. 问题描述

我创建了一个名称为myTest的topic, 该topic有三个分区,在我的应用中spark streaming以direct方连接kafka, 但是发现只能消费一个分区的数据, 多次更换comsumer group依然如此。

## 2 环境配置

kafka集群环境

主机	IP	操作系统	kakfa
node1 afka_2.11-0.10.1.1	192.168.1.101		Centos 6.5
node2 afka_2.11-0.10.1.1	192.168.1.102		Centos 6.5
node3 afka_2.11-0.10.1.1	192.168.1.103		Centos 6.5

应用依赖: spark版本是2.1.1、kafka版本是0.10.1.1;

maven依赖配置如下

```
<dependency>
<groupId>org.apache.spark</groupId>
<artifactId>spark-streaming_2.11</artifactId>
<version>2.1.1</version>
</dependency>
<dependency>
<groupId>org.apache.spark</groupId>
<artifactId>spark-streaming-kafka-0-10_2.11</artifactId>
<version>$2.1.1</version>
</dependency>
<dependency>
<groupId>org.apache.kafka</groupId>
<artifactId>kafka_2.11</artifactId>
<version>0.10.1.1</version>
</dependency>
```

相关配置代码 (Java) 如下:

```
Map<String, Object> kafkaParams = new HashMap<>();
kafkaParams.put(ConsumerConfig.BOOTSTRAP_SERVERS_CONFIG, "192.168.1.101:9092,192.168.1.102:9092,192.168.1.103:9092");
kafkaParams.put(ConsumerConfig.GROUP_ID_CONFIG, "myGroup");
kafkaParams.put(ConsumerConfig.AUTO_OFFSET_RESET_CONFIG, "earliest");
kafkaParams.put(ConsumerConfig.ENABLE_AUTO_COMMIT_CONFIG, "true");
kafkaParams.put(ConsumerConfig.KEY_DESERIALIZER_CLASS_CONFIG, StringDeserializer.class);

kafkaParams.put(ConsumerConfig.VALUE_DESERIALIZER_CLASS_CONFIG, StringDeserializer.class);

Set<String> topics = new HashSet<String>(Arrays.asList("testTopic"));
```

```
JavaInputDStream<ConsumerRecord<Object, Object>> dStream = KafkaUtils.createDirectStream(
    jssc,
    LocationStrategies.PreferConsistent(),
    ConsumerStrategies.Subscribe(topics, kafkaParams));
```

### 3. 解决方案

经过查阅相关资料发现是由于Kafka 0.10.1.1的bug导致的。其实不仅仅是0.10.1.1，另外0.10.1.0和0.10.0.2也有这个问题。详细描述参考<https://issues.apache.org/jira/browse/KAFKA-4547>

最后我将kafka版本降到了0.10.0.1，解决了这个问题。

```
<dependency>
<groupId>org.apache.kafka</groupId>
<artifactId>kafka_2.11</artifactId>
<version>0.10.0.1</version>
</dependency>
```