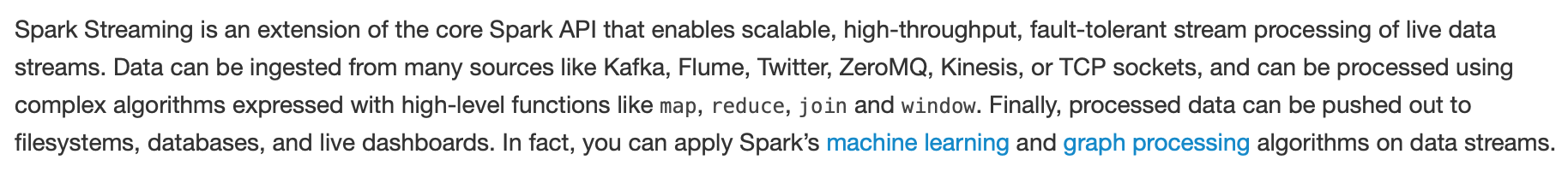
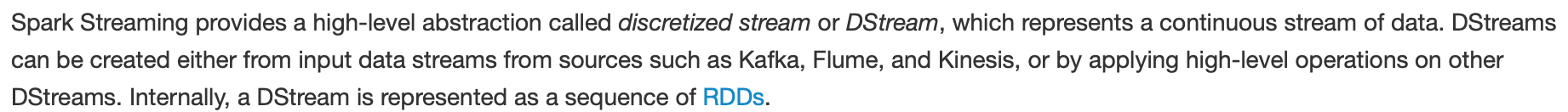
# Spark Streaming

* 概述





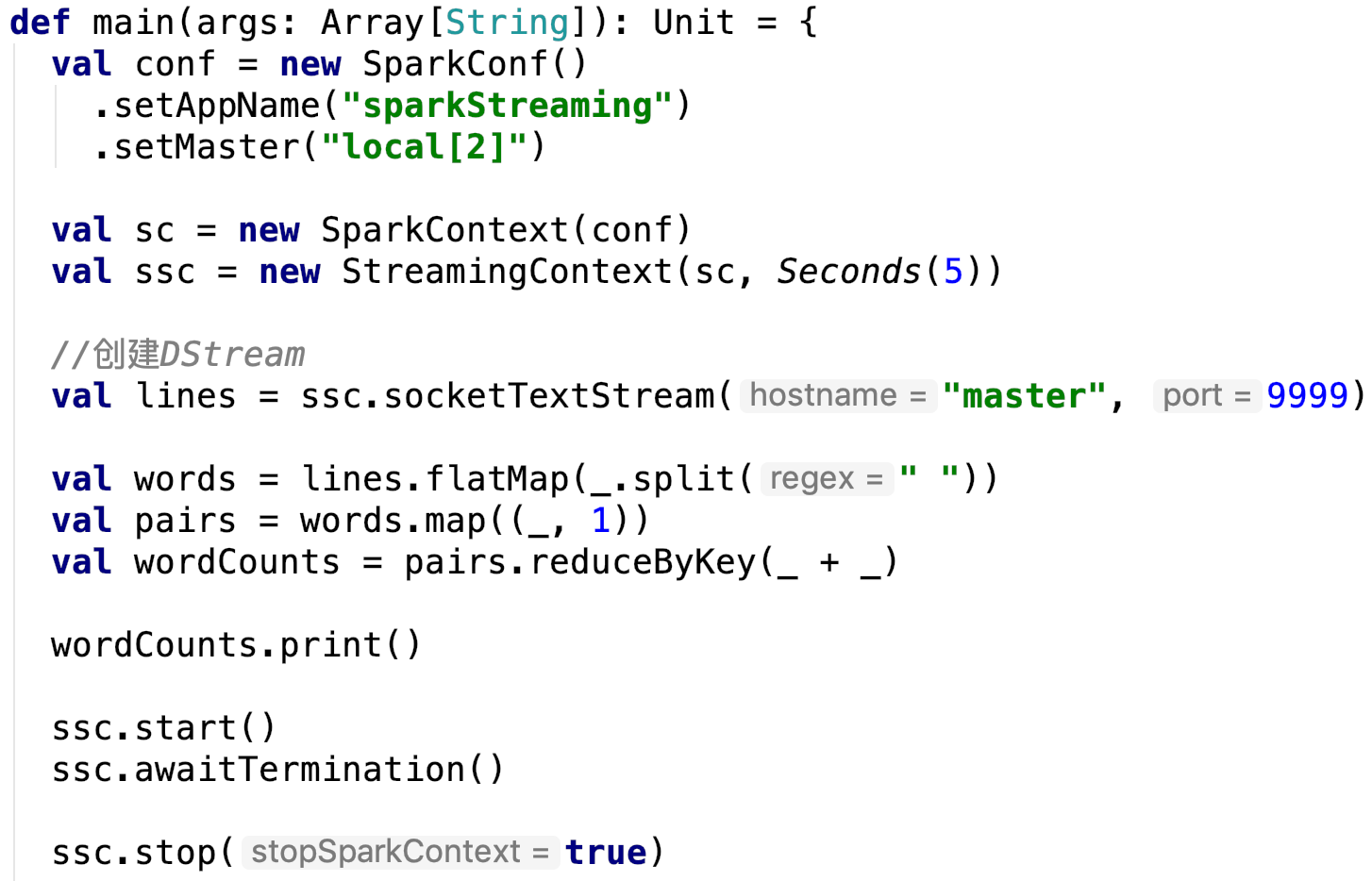




* wordCount程序开发

预备环境

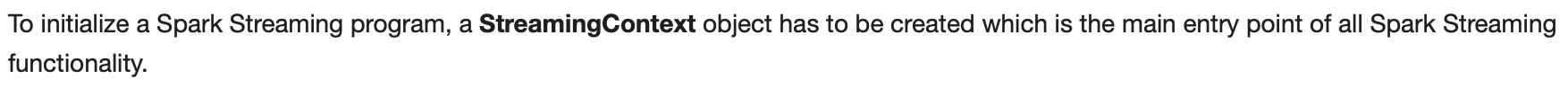
**yum install nc**

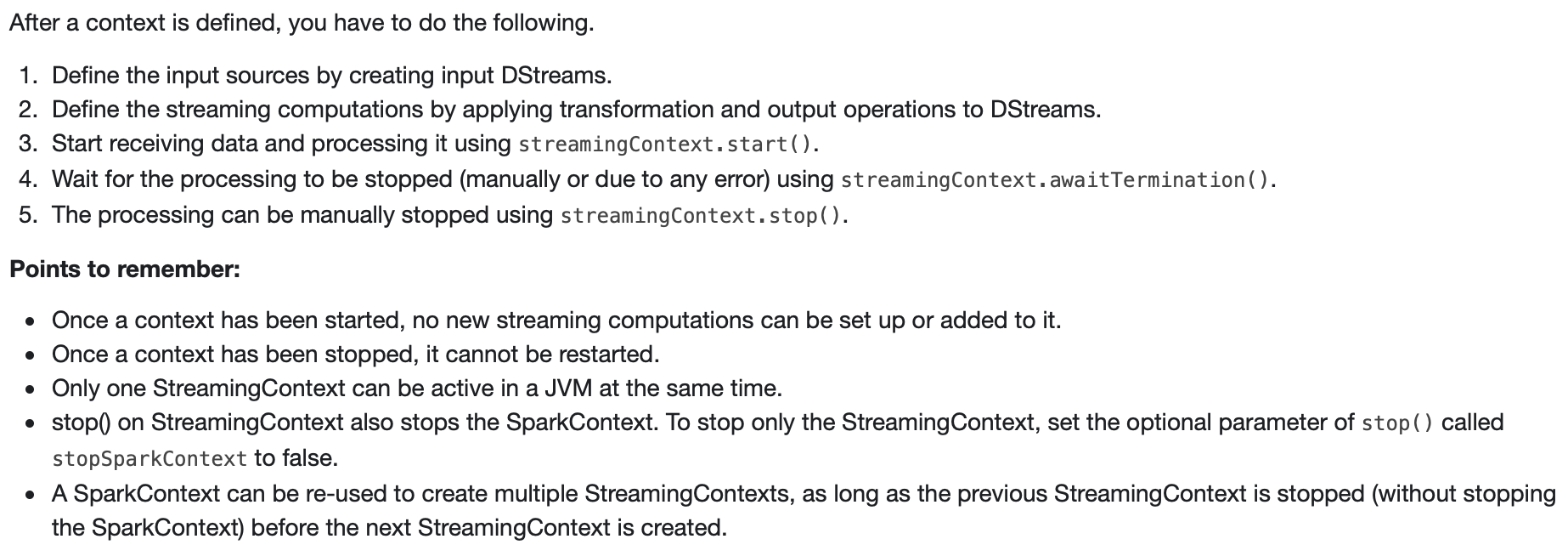


**注:local至少为local[2]，否则控制台没有输出**

* 基本概念

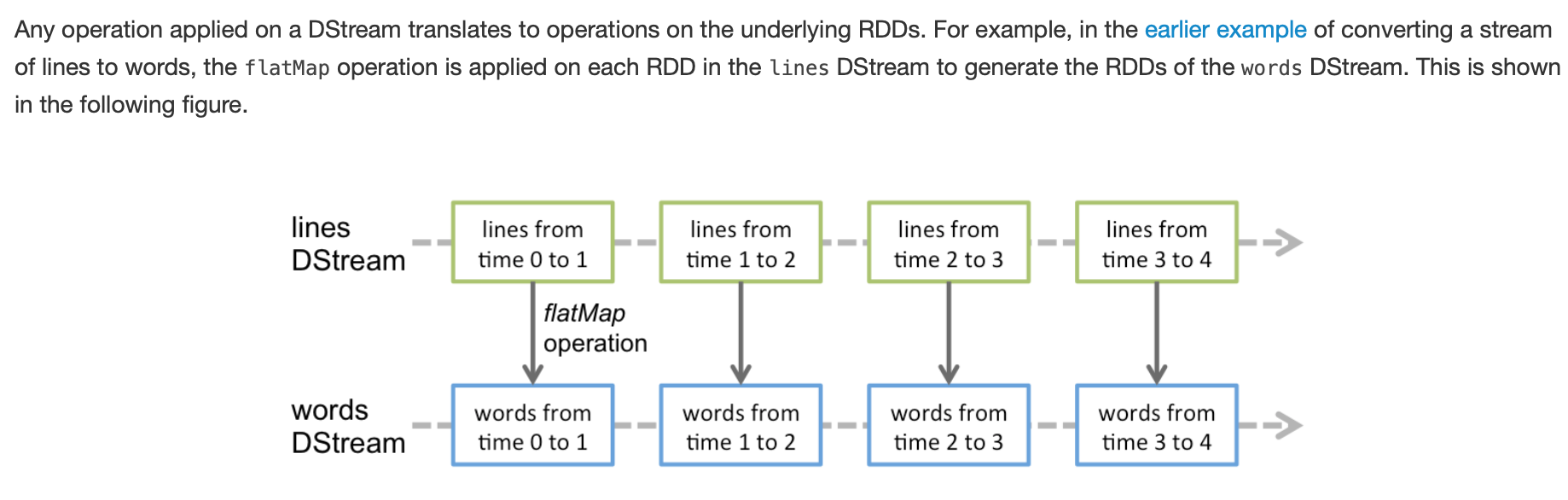
**StreamingContext**

****

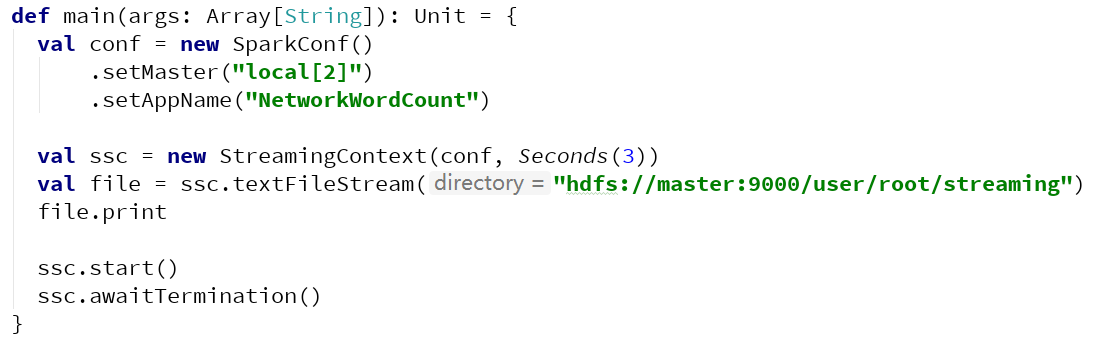
****

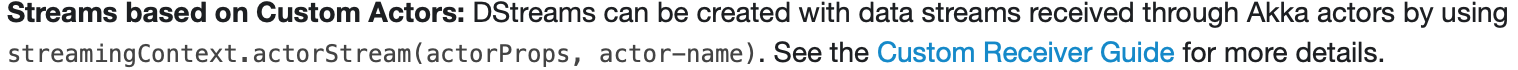
**DStream**

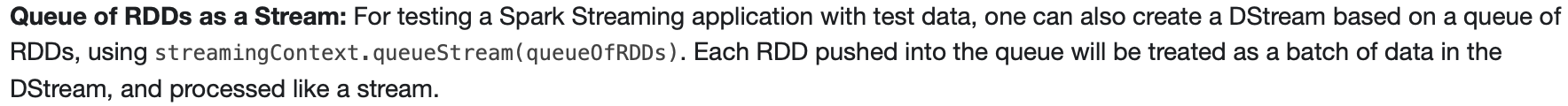




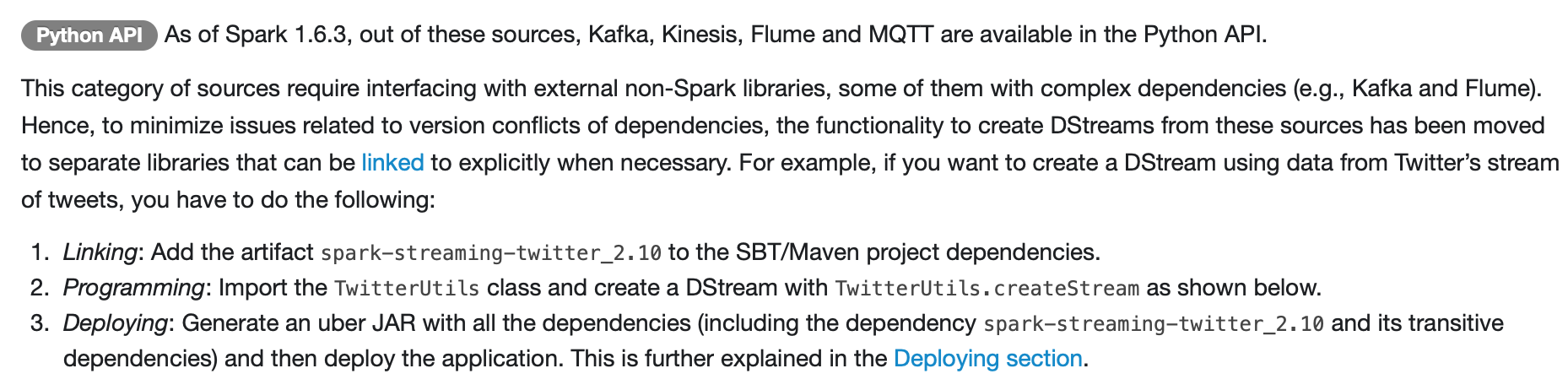
**基本数据源**





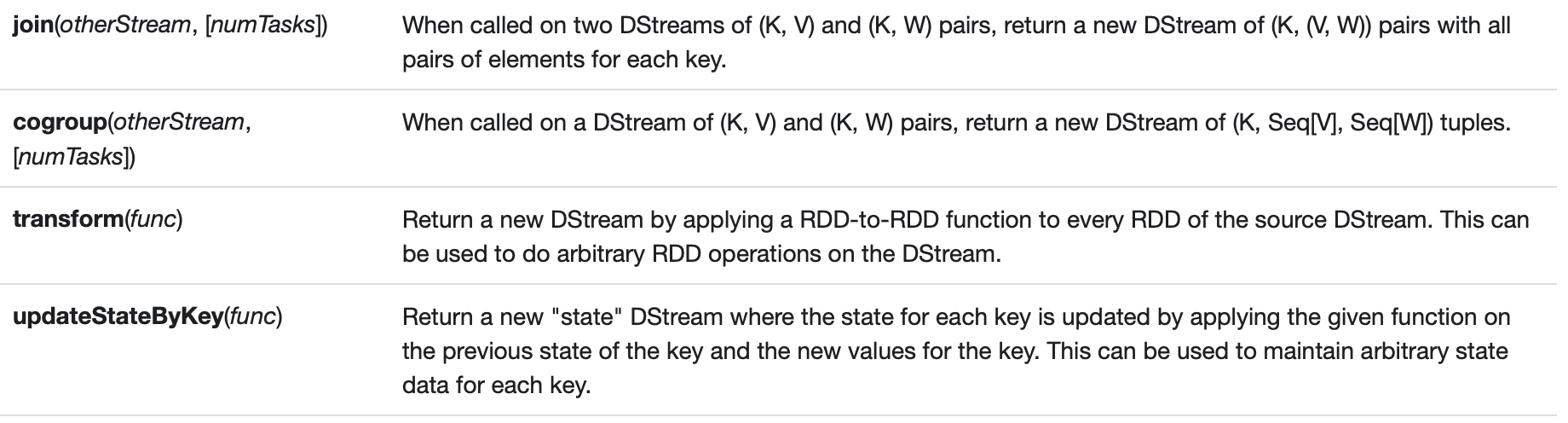
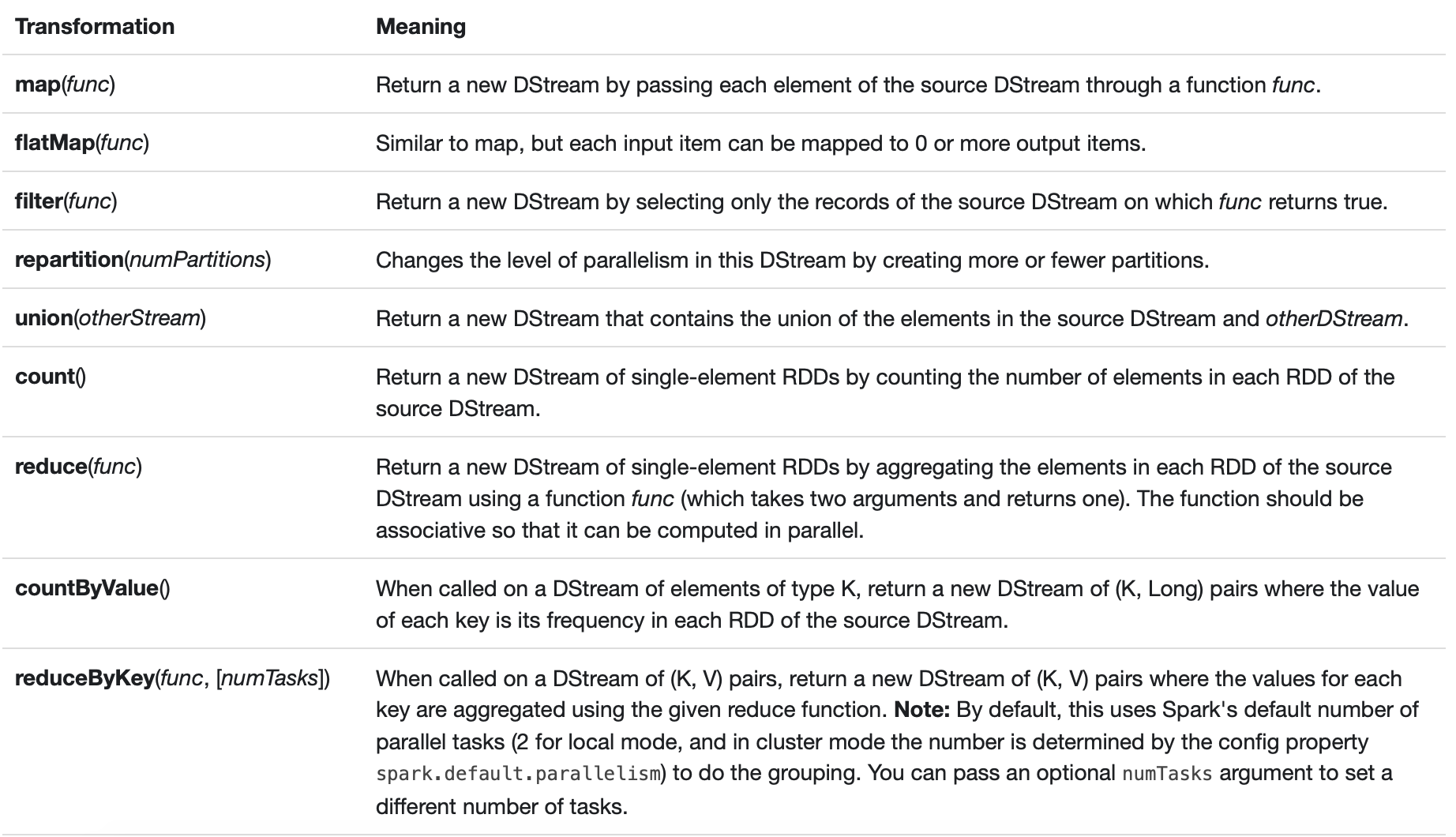
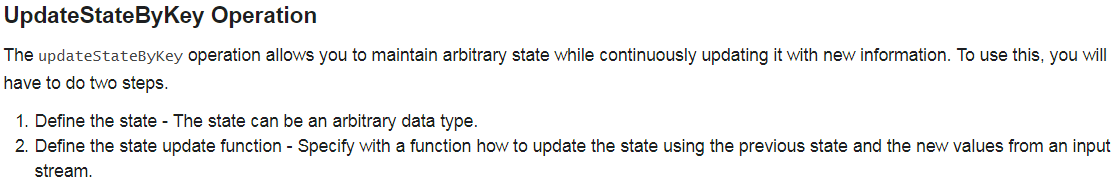
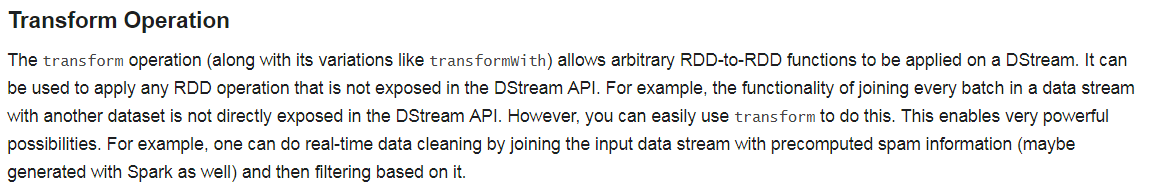


**高级数据源**

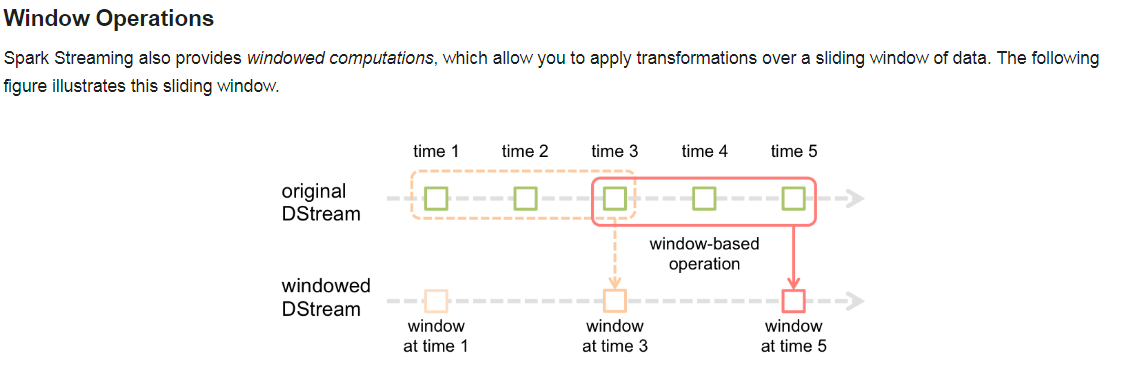
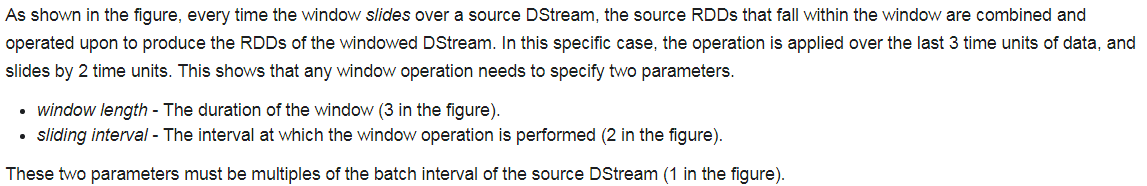


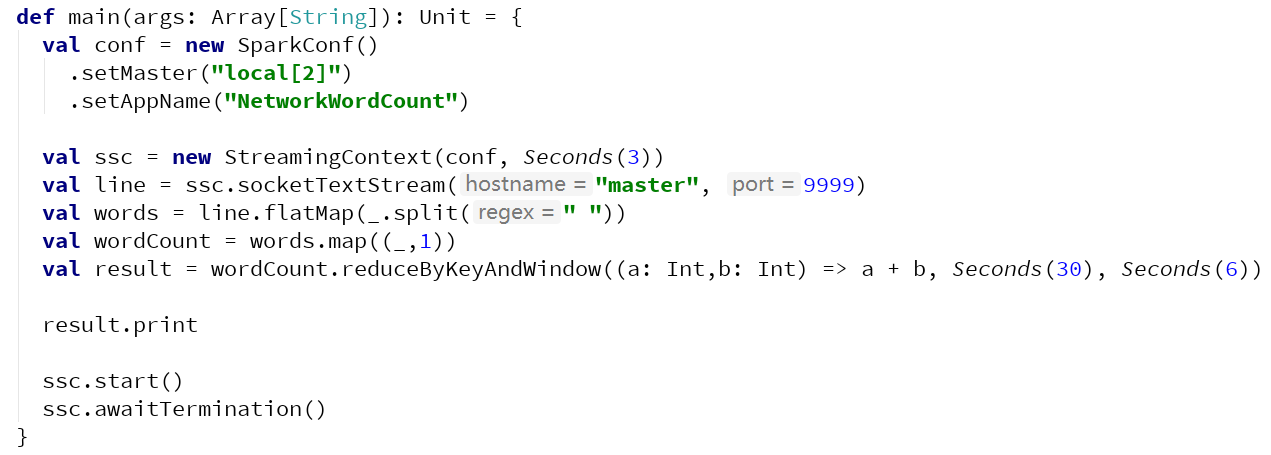
**DStream操作**

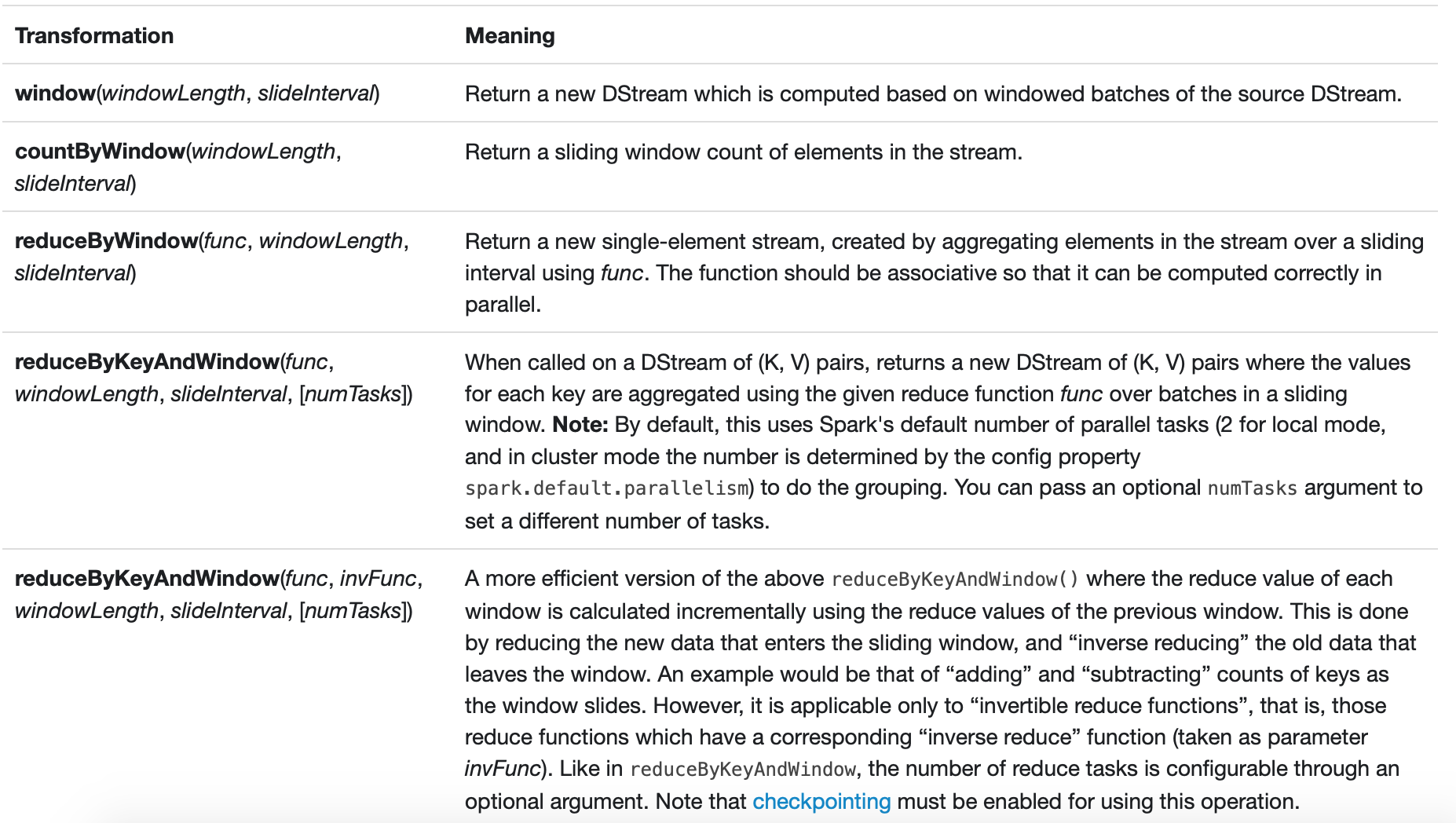
**Transformation操作**

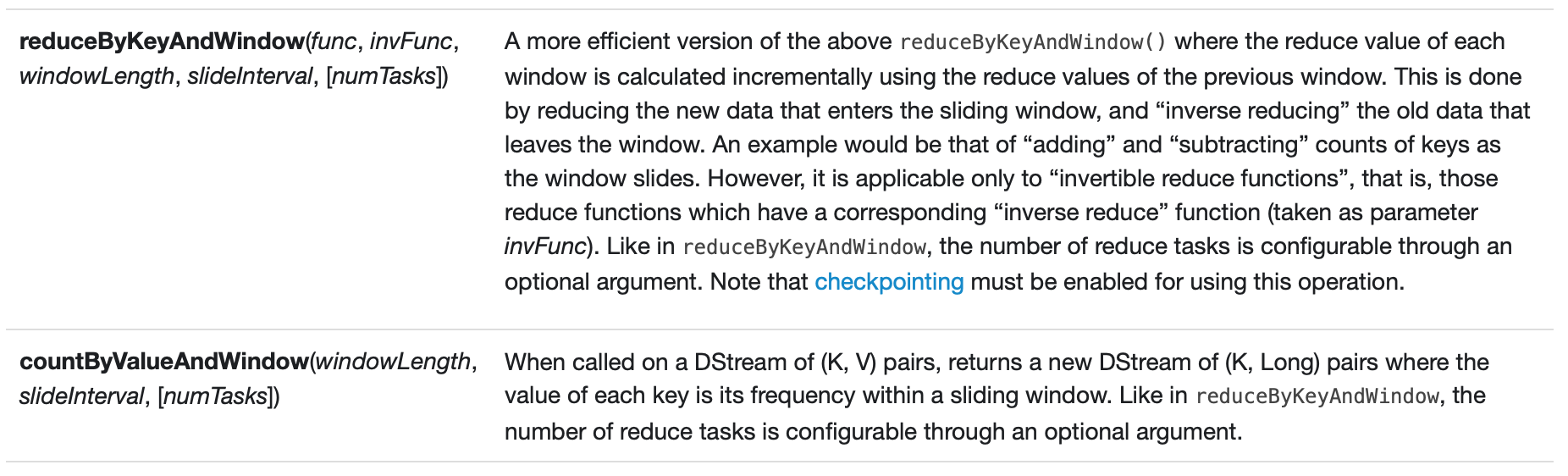
****  

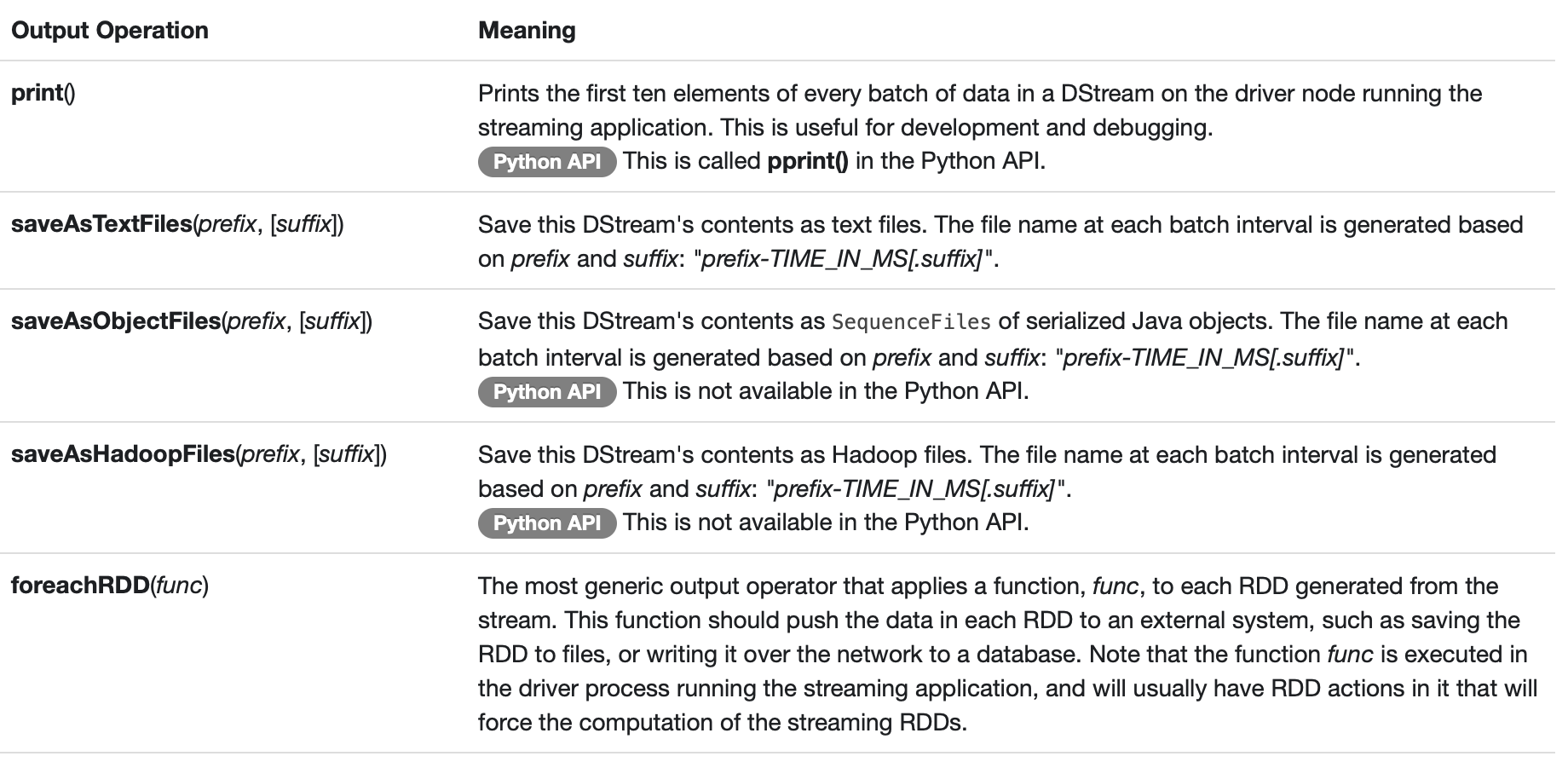


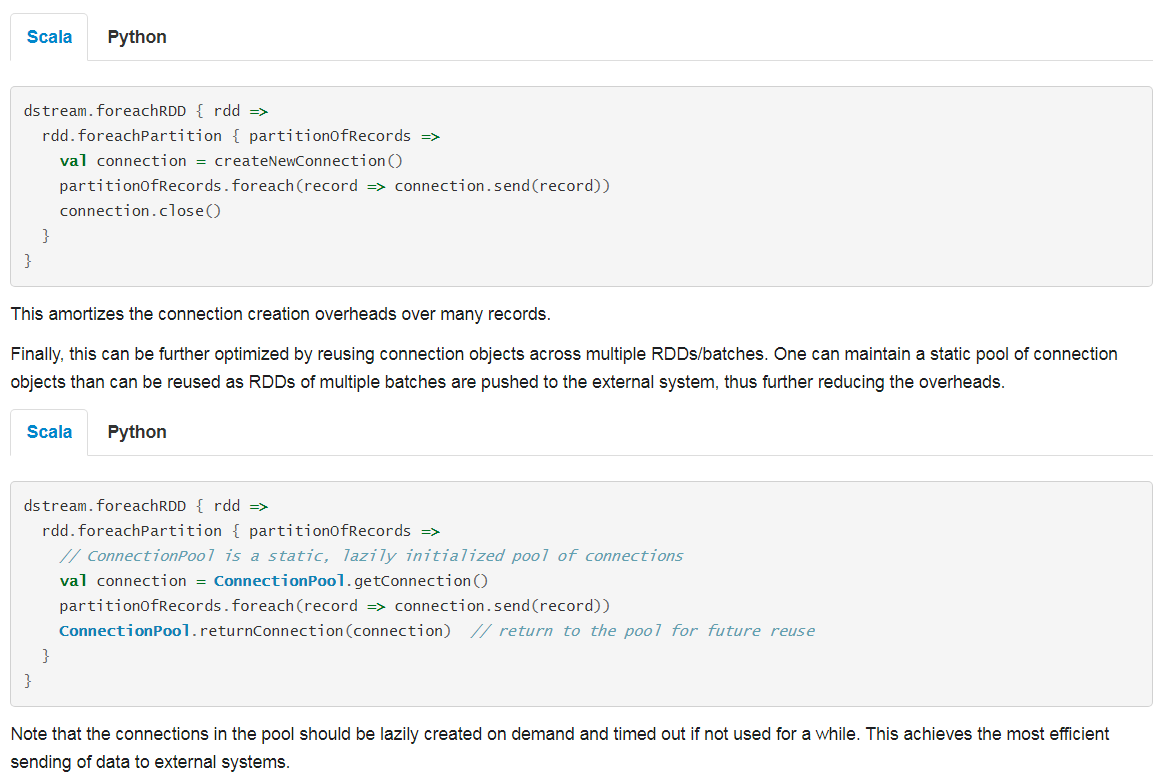
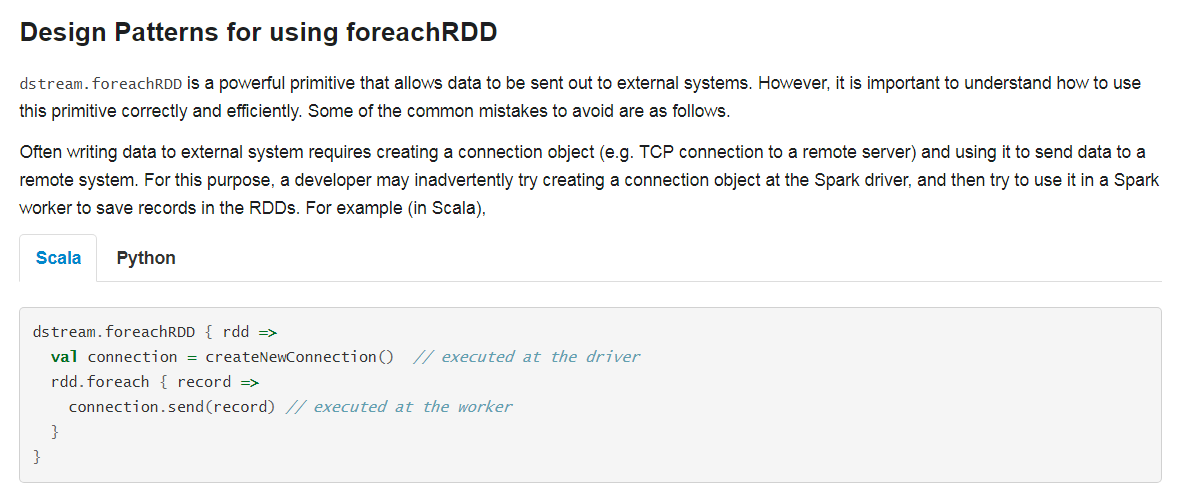
 



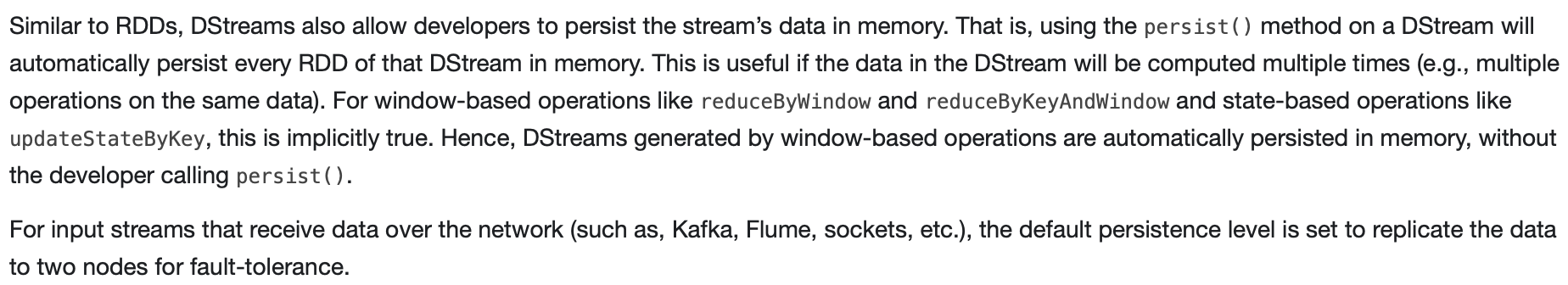
****

****

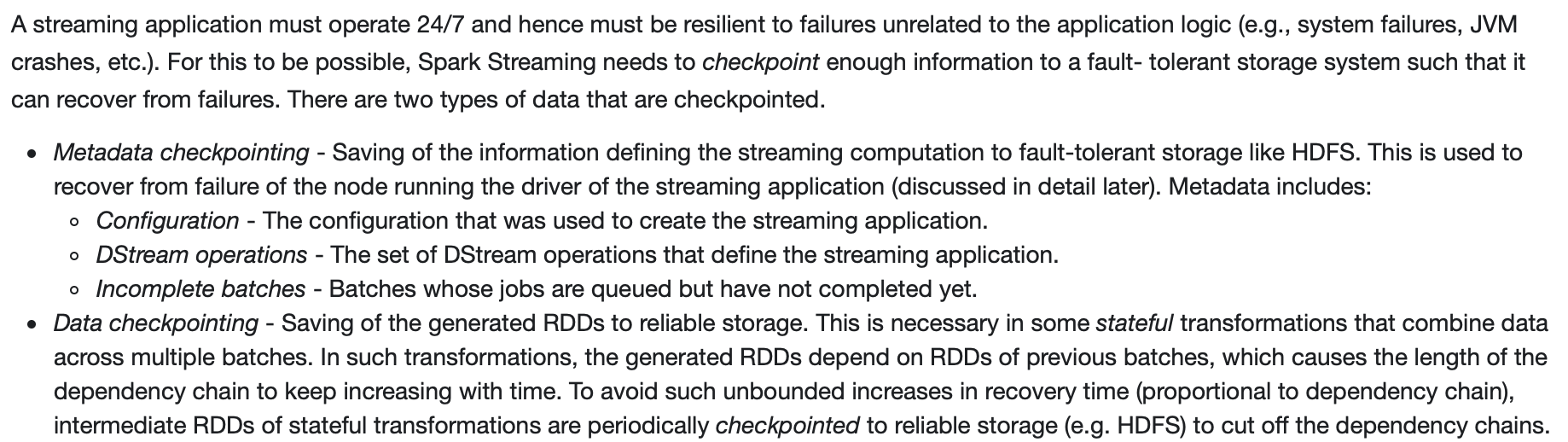
**Output操作**



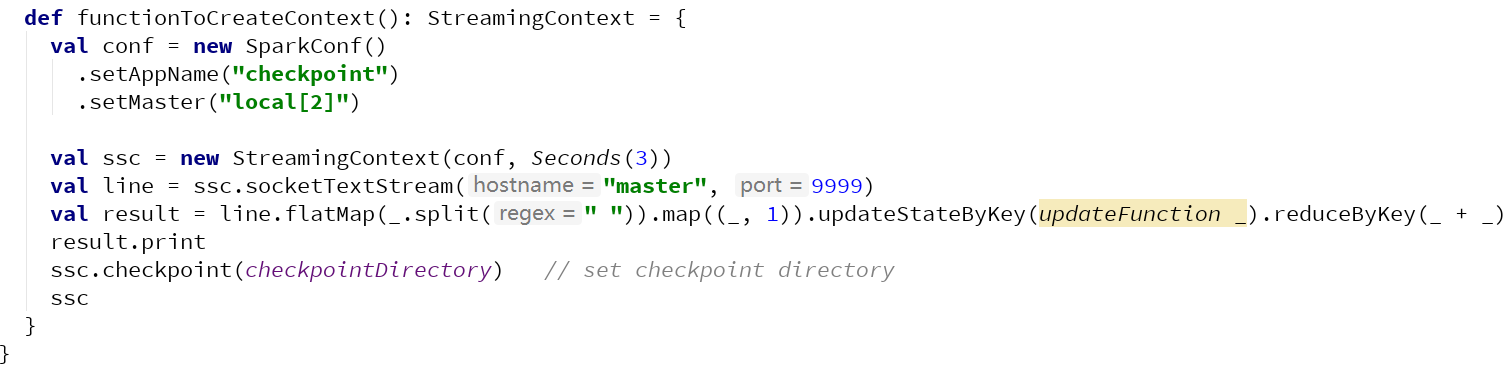
* 缓存/持久化

****

* 检查点

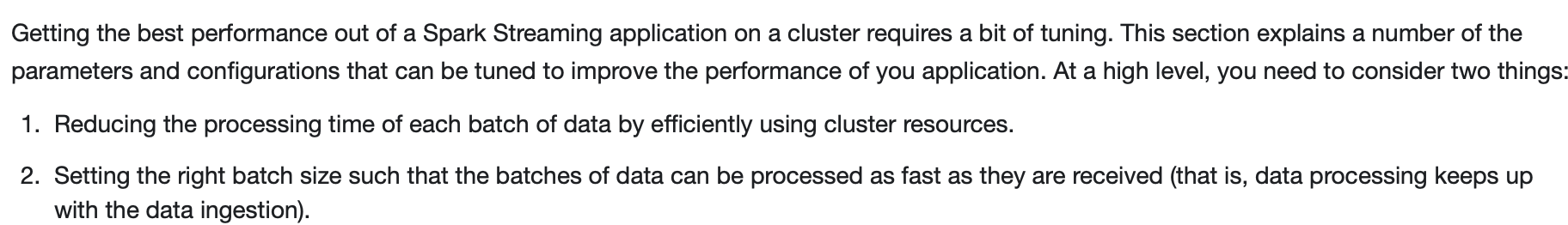






[**https://stackoverflow.com/questions/35090180/why-does-spark-throw-sparkexception-dstream-has-not-been-initialized-when-res**](https://stackoverflow.com/questions/35090180/why-does-spark-throw-sparkexception-dstream-has-not-been-initialized-when-res)

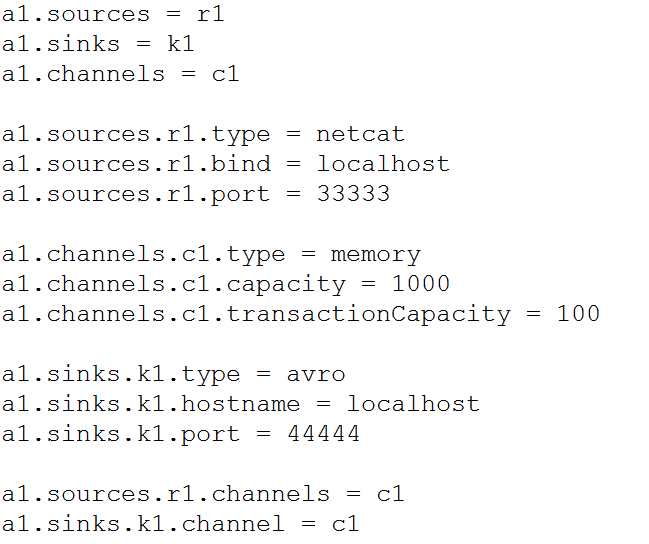
* 性能优化

****

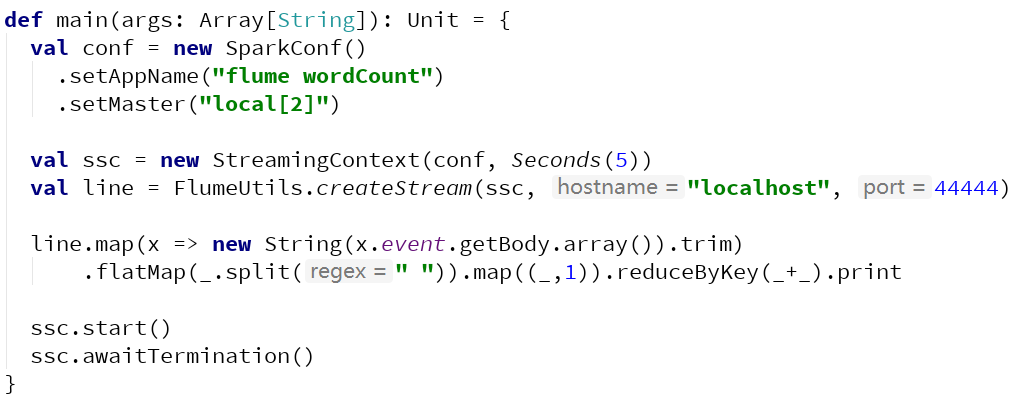
* Flume集成



**Flume配置**



**Spark Streaming应用开发**



**1.使用spark-submit启动spark streaming程序**

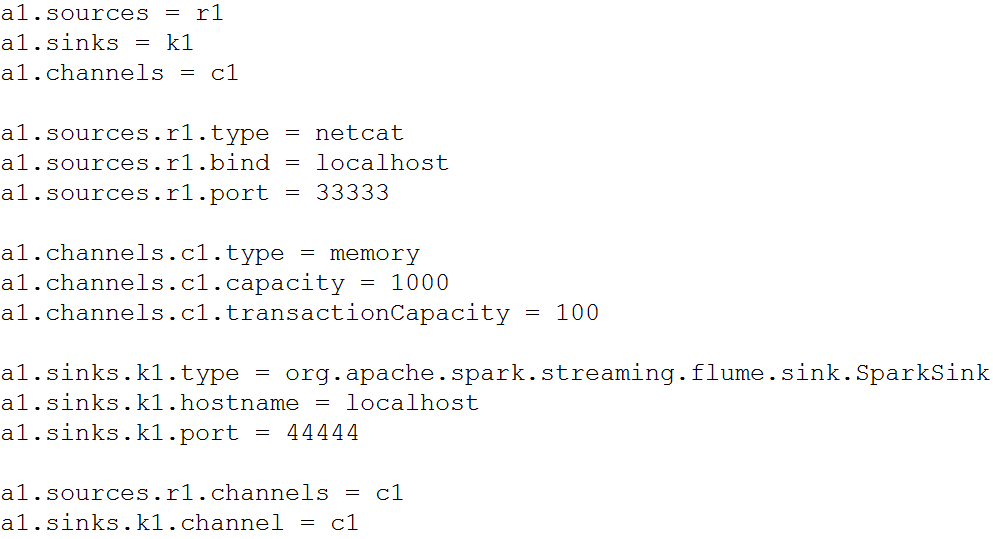
**2.启动flume程序**

**bin/flume-ng agent --name a1 --conf conf --conf-file conf/sparkstreaming\_flume.conf Dflume.root.logger=DEBUG,console**

**3.telnet localhost 33333进行测试**

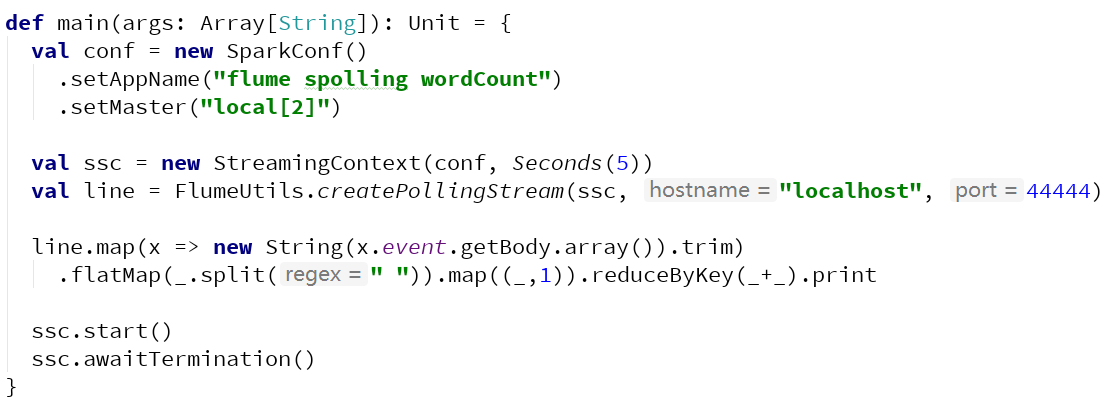


**Flume配置**



**注:在Flume与Spark目录下导入spark-streaming-flume-sink\_2.10-1.6.3.jar**

**Spark Streaming应用开发**



**1.启动flume程序**

**2.使用spark-submit启动spark streaming程序**

* Kafka集成



