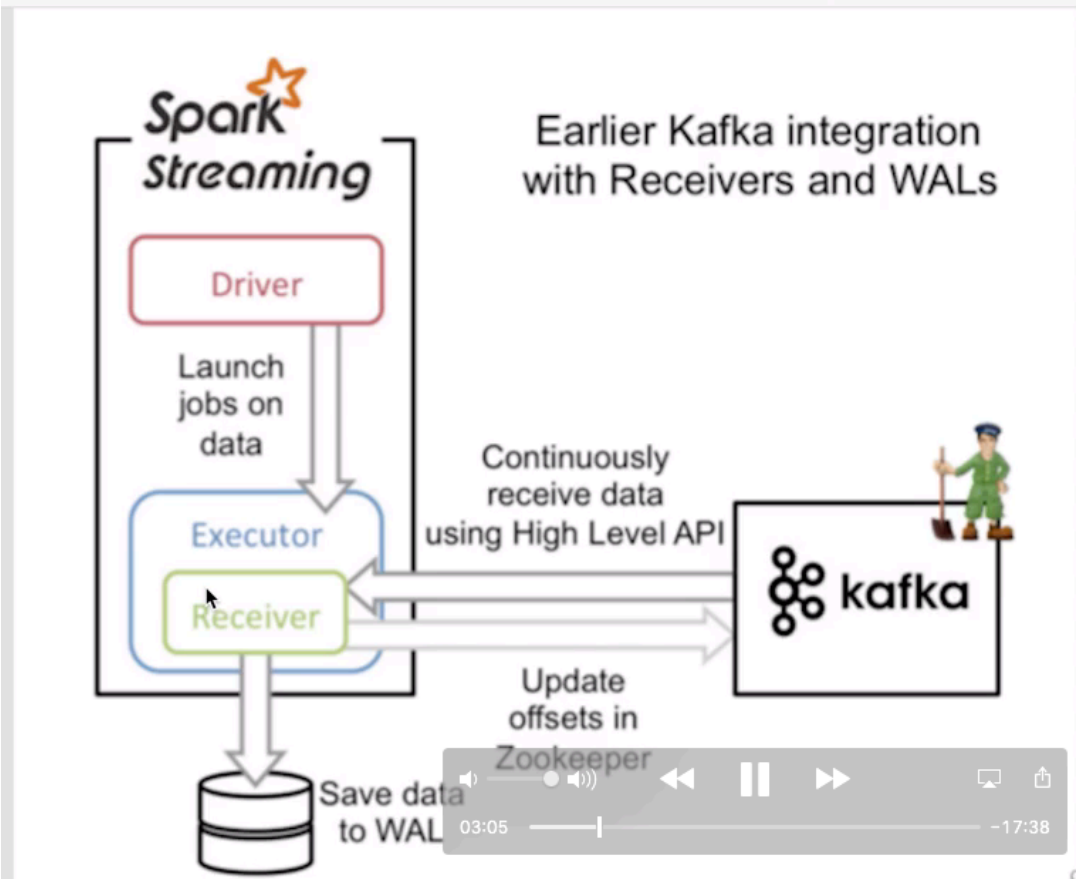
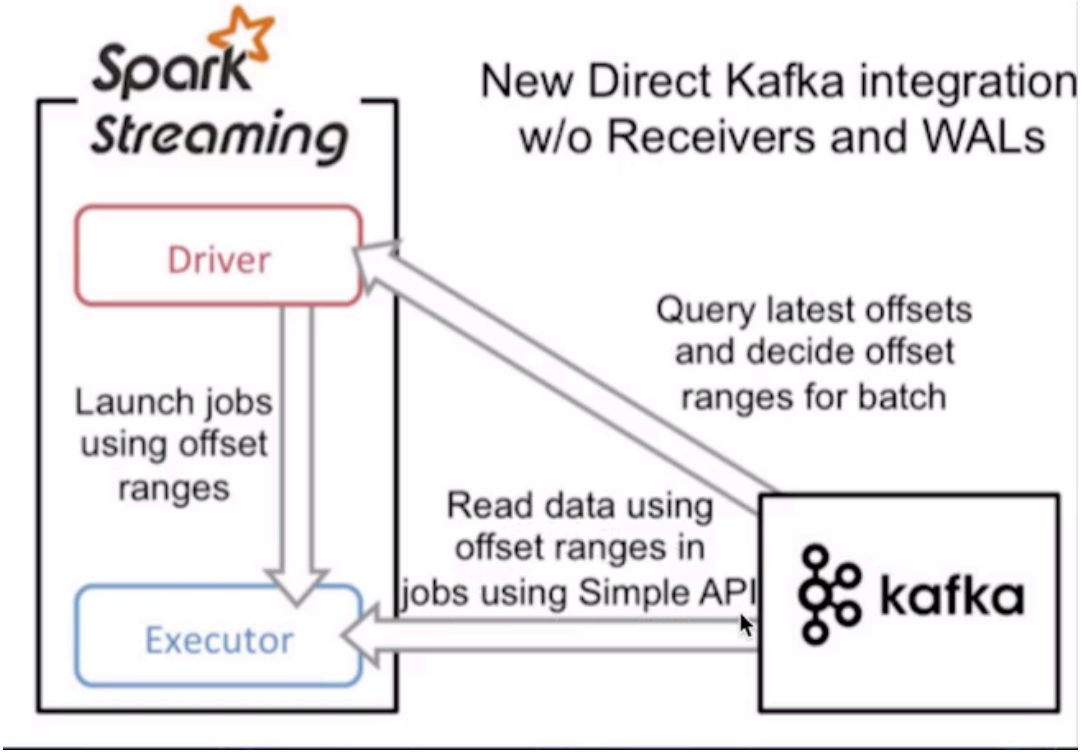
receiver 方式连接




```
12
13  SparkStraming的Receiver方式和直连方式有什么区别?
14      Receiver接收固定时间间隔的数据（放在内存中的），使用Kafka高级到API,自动维护偏移量，达到固定的时间才进行处理，效率低并且
        容易丢失数据
15      Direct直连方式，相当于直连连接到Kafka的分区上，使用Kafka底层的API,效率高，需要自己维护偏移量。
```

直接连接



1.Spark Streaming是一个基于Spark Core之上的实时计算框架，可以从很多数据源消费数据并对数据进行处理，
在Spark Streaing中有一个最基本的抽象叫DStream（代理），本质上就是一系列连续的RDD，DStream其实就是对RDD的封装
DStream可以任务是一个RDD的工厂，该DStream里面生产都是相同业务逻辑的RDD，只不过是RDD里面要读取数据的不相同

深入理解DStream:他是sparkStreaming中的一个最基本的抽象，代表了一下列连续的数据流，本质上是一系列连续的RDD，你对DStream进行操作，就是对RDD进行操作

DStream每隔一段时间生成一个RDD，你对DStream进行操作，本质上是对里面的对应时间的RDD进行操作

DSteam和DStream之间存在依赖关系，在一个固定的时间点，对个存在依赖关系的DSrteam对应的RDD也存在依赖关系，

每个一个固定的时间，其实生产了一个小的DAG，周期性的将生成的小DAG提交到集群中运行