



链滴

# Spark Streaming Kafka CreateDirectDStreaming 遇见的问题

作者: [daxin](#)

原文链接: <https://ld246.com/article/1482495030670>

来源网站: [链滴](#)

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

<p>问题 1: <br>

spark-submit 提交任务报错如下: <br>

分析: 起初我的 spark 集群是部署在 yarn 上, 所以在 spark-env 和 spark-default 下配置了 hadoop 相关参数。最后我想使用 spark standalone 模式跑程序, 就把 spark-env 和 spark-default 下的 hadoop 相关参数

注释掉了。之后提交程序提示: <br>

Exception in thread "main" java.net.ConnectException: Call From node1/192.168.88.130 to node1:9000 failed on connection exception: java.net.ConnectException: Connection refused; For more details see: <a href="https://ld246.com/forward?goto=http%3A%2F%2Fwiki.apache.org%2Fhadoop%2FConnectionRefused" target="\_blank" rel="nofollow ugc">http://wiki.apache.org/hadoop/ConnectionRefused</a> <br>

at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)<br>

异常: 就是说连接 hadoop hdfs 拒绝, 当时排错心想还是 spark conf 下有 hadoop 配置没注释完

。一遍遍检查最终也没找到, 最后怀疑是 linux 环境变量有 HADOOP\_CONF\_DIR 的配置, 结果使用 cho \$HADOOP\_CONF\_DIR<br>

果然存在, 在/etc/profile 中配置的, 注释掉解决问题。 <br>

总结: 尽量不要把局部应用环境变量配置在/etc/profile 中, 而是配置大数据框架的环境变量中。 </p>

<p>问题 2: <br>

博主是 Linux 单节点的伪分布式, 一个 Master 和一个 Worker, 并且 Master 和 Worker 在同一节点上。此时忽略了系统是跑在分布式的环境下的, 所以当时糊涂设置为本地

文件系统路径, 提交应用之后发现提示, 并且 CheckPoint 没有生效。 <br>

WARN SparkContext: Spark is not running in local mode, therefore the checkpoint directory must not be on

the local filesystem. Directory 'file:/home/daxineckpoint' appears to be on the local filesystem </p>

<p>spark 应用跑在集群模式下, checkpoint directory 是不可以设置在本地文件系统的, 因为程序在分布式环境下运行。 </p>

```
<p>package com.sparkstreaming.direct</p>
```

```
<p>import kafka.serializer.StringDecoder<br>
```

```
import org.apache.spark.streaming.kafka.KafkaUtils<br>
```

```
import org.apache.spark.{SparkConf, SparkContext}<br>
```

```
import org.apache.spark.streaming.{Seconds, StreamingContext}</p>
```

```
<p>/**</p>
```

```
<ul>
```

```
<li>Created by Dax1n on 2016/12/1.<br>
```

```
*/<br>
```

```
object DirectCreateDStream1 {<br>
```

```
val kafkaParams = Map[String, String](<br>
```

```
"metadata.broker.list" -> "node1:9092,node1:9093,node1:9094",<br>
```

```
"group.id" -> "onlyOneCk1")</li>
```

```
</ul>
```

```
<p>def main(args: Array[String]): Unit = {</p>
```

```
<pre> <code class="highlight-chroma"> <span class="highlight-line"> <span class="highlight-cl">val conf = new SparkConf()
```

```
</span> </span> <span class="highlight-line"> <span class="highlight-cl">conf.setAppName("LocalDirect").setMaster("local[2]")
```

```
</span> </span> <span class="highlight-line"> <span class="highlight-cl">val sc = new SparkContext(conf)
```

```
</span> </span> <span class="highlight-line"> <span class="highlight-cl">sc.setLogLevel("WARN")
```

```
</span> </span> <span class="highlight-line"> <span class="highlight-cl">
```

```
</span> </span> <span class="highlight-line"> <span class="highlight-cl">
```

```
</span> </span> <span class="highlight-line"> <span class="highlight-cl">
```

```

def createStreamingContext():StreamingContext={
  val ssc = new StreamingContext(sc, Seconds(2))
  ssc.checkpoint("C:\\streamingcheckpoint1")
  val dStream = KafkaUtils.createDirectStream[String,String,StringDecoder,StringDecoder](ssc,kafkaParams,Set("orderNumOnlyOne1"))
  dStream.map{
    x=>
      x._1+" - "+x._2
  }
  dStream1.print()
  ssc
}

```

// 重重注意：对于 Spark 的 Transform 和 Action 都要写在 getOrCreate 的 createStreamingContext 函数中，否则报错!!!，此处更多技巧看官方文档

官网地址：<https://ld246.com/forward?goto=http%3A%2F%2Fspark.apache.org%2Fdocs%2Flatest%2Fstreaming-programming-guide.html> 的 Checkpointin 章节

//

```

val ssc = StreamingContext.getOrCreate("C:\\streamingcheckpoint1",createStreamingContext_

```

//错误信息：

```

16/12/01 09:04:38 ERROR streaming.StreamingContext: Error starting the context, marking it
as stopped

```

```

org.apache.spark.SparkException: org.apache.spark.streaming.dstream.MappedDStream@4c2a67cc
has not been initialized

```

```


```

ssc.start()

```


```

```


```

ssc.awaitTermination()

```


```

```


```

</p>

</p>

如果把 Spark 的 Transform 和 Action 写在创建 CreateStreamContext 函数外面会报如下错：

```

16/12/01 09:04:38 ERROR streaming.StreamingContext: Error starting the context, marking
it as stopped

```

```

org.apache.spark.SparkException: org.apache.spark.streaming.dstream.MappedDStream@4c2a67cc
has not been initialized

```

错误代码：

```

package com.sparkstreaming.direct

```

```

import kafka.serializer.StringDecoder

```

```

import org.apache.spark.streaming.kafka.KafkaUtils<br>
import org.apache.spark.{SparkConf, SparkContext}<br>
import org.apache.spark.streaming.{Seconds, StreamingContext}</p>
<p>/**</p>
<ul>
<li>Created by Dax1n on 2016/12/1.<br>
*/<br>
object DirectCreateDstream1 {<br>
val kafkaParams = Map[String, String](<br>
"metadata.broker.list" -&gt; "node1:9092,node1:9093,node1:9094",<br>
"group.id" -&gt; "onlyOneCk1")</li>
</ul>
<p>def main(args: Array[String]): Unit = {</p>
<pre> <code class="highlight-chroma"> <span class="highlight-line"> <span class="highlight-cl"> val conf = new SparkConf()
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> conf.setAppName
"LocalDirect").setMaster("local[2]")
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> val sc = new Spar
Context(conf)
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> sc.setLogLevel("
ARN")
</span> </span> <span class="highlight-line"> <span class="highlight-cl">
</span> </span> <span class="highlight-line"> <span class="highlight-cl">
</span> </span> <span class="highlight-line"> <span class="highlight-cl">
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> def createStreami
gContext():StreamingContext={
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> val ssc = new Str
amingContext(sc, Seconds(2))
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> ssc.checkpoint("C
\\streamingcheckpoint1")
</span> </span> <span class="highlight-line"> <span class="highlight-cl">
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> ssc
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> }
</span> </span> </code> </pre>
<p>//错误写法: Transform 和 Action 写在创建 CreateStreamContext 函数外面<br>
val ssc = StreamingContext.getOrCreate("C:\streamingcheckpoint1",createStreamingContext _
</p>
<pre> <code class="highlight-chroma"> <span class="highlight-line"> <span class="highlight-cl"> val dStream = KafkaUtils.createDirectStream[String,String,StringDecoder,StringDecoder
(ssc,kafkaParams,Set("orderNumOnlyOne1"))
</span> </span> <span class="highlight-line"> <span class="highlight-cl">
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> val dStream1 =
Stream.map{
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> x=&gt;
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> x._1+" - "+x._2
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> }
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> dStream1.print()
</span> </span> <span class="highlight-line"> <span class="highlight-cl">
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> ssc.start()
</span> </span> <span class="highlight-line"> <span class="highlight-cl"> ssc.awaitTerminat
on()
</span> </span> </code> </pre>
<p></p>

```

<p>}</p>

<p><em><strong>我的 CSDN:<a href="https://ld246.com/forward?goto=http%3A%2F%2Fblog.csdn.net%2Fdax1n%2Farticle%2Fdetails%2F53425668" target="\_blank" rel="nofollow ugc">http://blog.csdn.net/dax1n/article/details/53425668</a></strong></em></p>