

Spooling Directory Source 是使用 `tail -F` 的 Exec Source（将稍后在本章讨论）的一种好的替换方案，因为这类 Source 能保证数据传送，通常比使用 `tail -F` 的 Exec Source 更加可靠。唯一的缺点是数据不是实时跟踪的，并且只能在文件关闭或移入到相关目录时才能读取文件。文件一旦被 Source 完全使用完且所有的事件被成功写入 Source 的 Channel 中，Source 就可以基于配置重命名文件或删除文件。当文件被重命名，Source 只是给文件名添加一个后缀，而不是完全改变它。这个后缀也是可以配置的。

Spooling Directory Source 使用追踪器持久化到磁盘，以定位每个文件在哪个位置成功将事件写入 Channel，这样如果 Agent 或机器失败和重启，Source 就能从这个位置开始读取数据。这也能保证 Source 实时定位任一位置的处理，并且保证当 Source 重启时，处理是接着上次处理的位置。这正是 Source 不允许文件名重用的原因之一。

表 3-6 展示了 Spooling Directory Source 接受的各种配置参数。

表3-6 Spooling Directory Source配置

参数	默认值	描述
type	-	Spooling Directory Source 的别名是 <code>spooldir</code> ，FQCN 是 <code>org.apache.flume.source.SpoolDirectorySource</code>
spoolDir	-	监控和读取文件的目录。该目录的子目录不会被扫描
batchSize	100	每个事件写入 Channel 的最大事件数量
ignorePattern	^\$	文件名称匹配这个正则表达式被忽略和数据也不会从中读取
deletePolicy	never	如果要删除使用完的文件——该参数的值为 <code>never</code> 或 <code>immediate</code>
fileSuffix	.COMPLETED	被使用完的文件使用的后缀。如果这是扩展“.”是必需的
fileHeader	false	如果设置为 <code>true</code> ，文件名会被添加到事件的 header
fileHeaderKey	file	如果文件名被添加到 header，此参数为在 header 中使用的密钥
trackerDir	.flumespool	Spooling Directory Source 存储元数据的目录，用来 source 中断时来重启 Source
deserializer	line	Builder 类的别名或完整类别名称，用来创建读取自定义格式数据的反序列化器。下节解释了如何写反序列化器
deserializer.*	-	任何传递给反序列化器的参数
inputCharset	UTF-8	当反序列化器调用 <code>readChar</code> 方法时使用的编码