Spooling Directory Source 的类型是 spooldir。在前面提及过，Source 读取指定目录中的所有文件并且逐个处理它们。处理目录的完整路径应该通过 spoolDir 参数来传递。由于性能的原因，Source 批次写事件，批处理的最大数量通过 batchSize 参数定义。Source 尝试尽可能多地从文件中读取事件，直到达到指定的批量大小。如果文件中有更少的可用事件，一旦文件中所有事件都读取完，它就尽快提交事务。

有时，写文件到相同的目录中，实际上可能不包含数据，如元数据文件。为了避免使用此类文件，即不包含有效数据的文件，可以通过使用 ignorePattern 参数指定 ignore 模式。 52
这个参数需要一个正则表达式，任何匹配正则表达式文件名的文件将被忽略。

前面提及过，文件一旦完全使用完，Flume 就能重命名或删除文件。如果要立即删除文件，设置 deletePolicy 参数为 immediate 即可。如果 deletePolicy 设置为 never（默认值），文件一旦使用完就重命名，用 fileSuffix 参数指定的后缀追加到原始文件名的后面。任何针对已完成的文件使用的该后缀将被忽略，所以要谨慎，不要使用这样的文件后缀，即写入该目录的新文件的后缀。

当文件被处理完并且从文件生成事件时，通常有利于处理系统知道事件来自哪些文件（例如，在搜索 UI 中展示属于堆栈跟踪的文件名）。通过设置 fileHeader 参数为 true 可以包含完整路径和文件名。使用 fileHeaderKey 参数可以设置 header 中使用的秘钥（默认值为 file）。

Spooling Directory Source 能够从它中断的位置恢复，所以能避免重复消耗文件中的数据。这使得当 Source 启动的时候，持久化文件处理和读取的信息到磁盘是可以实现的。信息持久化到追踪目录中，追踪目录一直在 Source 监控的目录中。追踪目录默认的名字是 .flumespool，这可以通过 trackerDir 参数来改变。要注意的是，目录内创建的子目录将会被读取，trackerDir 参数的值被用作 Source 监控目录的相对路径。一旦追踪目录的名字设置好，如果这个参数的值发生改变（甚至关闭 Flume 之后），Source 将不能再定位文件处理的位置，可能在开始的位置结束处理导致重复。所以，该参数一旦设置就不能修改。

例 3-7 展示了一个例子，关于使用 Spooling Directory Source 从目录中以每批次 250 个事件的速度读取数据到磁盘。当文件完全使用完的时候 Source 会尽快删除文件。通过 UsingFlumeFiles 的密钥也能向 header 插入文件名。

例3-7 Spooling Directory Source配置示例

```
agent.sources = spool
agent.channels = memChannel

agent.sources.spool.type = spooldir
```