

时间段结束时写入缓存的事件，这可以通过 `batchTimeout` 参数设置。如果 `batchSize` 和 `batchTimeout` 参数都设置好了，一旦达到批处理大小或批处理超时，就批量写入到 Channel 中。

Exec Source 可以设置为在一个单独的 shell 中运行配置命令，这可能不同于 Flume 处理所运行的方式。要在不同的 shell 中运行该处理，传递可执行 shell 的完整路径给 shell 参数。如果命令需要 shell 突出通配符或管道的替换功能，那么 shell 参数必须设置，因为 Flume 不会执行替换。

64

下面是 Exec Source 配置文件的示例，使用 shell 参数来运行一个完整的命令：

```
agent.sources = execSrc
agent.channels = memChannel

agent.sources.execSrc.type = exec
agent.sources.execSrc.shell = /bin/bash -c
agent.sources.execSrc.command = tail -F /var/log/flume/flume.log | grep
"error:"

agent.sources.execSrc.channels = memChannel

# Initializes a memory channel with default configuration
// 使用默认配置初始化 memory channel
agent.channels.memChannel.type = memory
```



Exec Source 丢失数据的可能性

Exec Source 是异步 Source 的一个例子，即如果有失败可能通知不到数据生产者。因此，重新启动 Agent 或机器会导致数据丢失，同上面解释的一样。

Exec Source 在 Flume 中最常用来追踪文件。利用 `tail -F` 命令使用 Exec Source 追踪文件，将近乎实时地将数据放入 Flume，但存在数据丢失的风险。如果 Flume Agent 死亡或机器重启，Exec Source 将在它启动的时候运行该命令；在这种情况下，它将运行 `tail -F <file_name>` 命令。因为 `tail` 命令只会拉取新数据写入到文件中的新数据，任何在 Agent 死亡和 Source 启动期间的写入文件的数据都会丢失。由于这个原因，建议使用本章前面所讨论的 Spooling Directory Source 处理写入文件的数据。尽管更严格一些，但因为它追踪的是正在从文件中读取的数据，所以 Source 不会丢失数据。

即使使用其他一些命令，Exec Source 在将事件写入 Channel 之前，也能缓冲和 batch 大小尽可能多的事件。如果 Agent 或机器重启，这些事件也可能会在批处理超时或达到 batch 最大容量之前丢失。