

# Exec Source

Exec Source 执行用户配置的命令，且基于命令的标准输出来生成事件。它还可以从命令中读取错误流，将事件转换为 Flume 事件，并将它们写入 Channel。Source 希望命令不断生产数据，并且吸收其输出和错误流。只要命令开始运行，Source 就要不停地运行和处理，不断读取处理的输出流（如果进行了配置，还有错误流）。

接着，输出流中的每一行将被编码为字节数组。要使用的编码是可配置的，默认为 UTF-8。然后每个字节数组用作 Flume 事件的 body。由于性能原因，Source 批量预配置事件的数量（或直到超时），并把它们写入 Channel。如果 Channel 已满，Source 可以配置成停止读取流程的输出和错误流（因此暂时阻塞流程），或放弃当前的批处理，继续读取输出和错误流（因此可以让处理继续）。表 3-12 举例说明了 Exec Source 的配置参数。

表3-12 Exec Source配置

配置参数	默认值	描述
type	-	Exec Source 的别名是 exec，FQCN 是 org.apache.flume.source.ExecSource（大小写敏感）
command	-	Source 运行的命令
restart	false	如果设置为 true，当流程死亡时 Source 将重启流程
restartThrottle	10000	重启流程前要等待的毫秒值时间。如果 restart 参数值为 false 或没有设置，则该参数无影响
logStdErr	false	如果设置为 true，流程的错误流也会被读取并转换为 Flume 事件
batchSize	20	配置的在写入 Channel 之前事件的最大数量
batchTimeout	3000	配置的在将缓存事件写入 Channel 前要等待的毫秒值时间
charset	UTF-8	编码输出和错误流为 Flume 事件要使用的字符集
shell	-	用于运行该命令的 shell 或命令处理器

通过 command 参数传递要运行的命令。通过设置 restart 参数为 true，Exec Source 可以配置为随着命令的开始重启过程。为了确保有足够的时间区别命令的执行，可以设置 restartThrottle 参数。这个过程一旦死亡，Source 将在命令再次运行之前等待这么长的时间间隔。

当与 File Channel 同时使用时，Exec Source 通过缓存数据来确保好的性能。在第 4 章“file-channel”一节中会提到，当每个事务的事件数量很大时，Channel 有更好的性能。Exec Source 中的 batchSize 参数控制了批处理的大小。Source 也能配置在配置的