## Sink 的生命周期

Sink 使用标准的 Flume 配置系统进行配置。每个 Agent 可以没有 Sink 或若干 Sink。每个 Sink 只能从一个 Channel 中读取事件。如果 Sink 没有配置 Channel,那么 Sink 就会从 Agent 中被移除。配置系统保证:

- 每个 Sink 至少有一个正确配置的 Channel 连接它。
- 每个 Sink 有一个定义的 type 参数。
- Sink 是在 Agent 中 Sink 活跃列表中的。

类似于 Source,每个 Sink(如果配置正确)被 ConfigurationProvider 实例化然后被配置。配置错误和 configure 方法抛出异常的 Sink 会从 Agent 中被移除。Sink 一旦配置好,就会被配置系统启动。Flume 可以聚合线程到 Sink 组,这个我们将在后面讨论。每个 Sink 组可以包含一个或多个 Sink。如果一个 Sink 没有定义 Sink 组,那么该 Sink 可以被认为是在一个组内,且该 Sink 是组内的唯一成员。

每个 Sink 运行器运行一个 Sink 组,Sink 运行器是一个针对 Sink 组循环调用 process 方法的线程,依次转发调用到组里其中一个 Sink 的 process 方法。如果 Sink 可以立即处理更多的事件,通常是当前调用至少处理了一个事件的时候,process 方法就返回 Status.READY。如果 Sink 不能处理任何事件,Sink 返回 Status.BACKOFF,表明运行线程在再次尝试之前应该回退一段时间。为 Sink 定义 Sink 组不是强制的。如果一个 Sink 是活跃的,但是不在 Sink 组里,那么它就获取了一个针对自身的 Sink 运行器,调用 Sink 的 process 方法。针对该 Sink 的回退语义和 Sink 组里的相同。我们将第 6 章 "Sink 组和 Sink 处理器"一节讨论什么时候使用 Sink 组和 Sink 处理器。

Sink 可以在配置中指定它们的 FQCN 或内置 Sink 的别名来指定。每个 Sink 必须有一个配置的 Channel。只有当 Channel 配置正确且启动成功,Sink 才会启动。表 5-1 列出了需要的参数。

## 表5-1 所有Sink必需的配置参数

95

| 参数      | 描述  |
|---------|---|
| type    | Sink 的类型。可以是 FQCN 或 Sink 的别名(这只针对 Flume 内置的 Sink)。该类必须在 Flume 的环境变量中。Flume 内置 Sink 的别名在相关部分将被提及 |
| channel | 读取事件的 Channel   |

类似于其他所有的组件(除了 Memory Channel), 当 Flume 被重新配置, Sink 的实例是不会被重用的。因此, Sink 不应该缓冲任何已被读取和提交的事件。这是因为提交事