

源 (Source)

Source 是负责接收数据到 Flume Agent 的组件。Source 可以从其他系统中接收数据，像 Java 消息服务 (JMS) 或者其他处理的输出结果。Source 也可以用于接收其他 Flume Agent 的 Sink 通过 RPC 发送的数据。Source 甚至可以生产数据。不夸张地说，Source 可以接收任何来源的数据。

数据源从外部系统或者其他 Agent (或者自身生产) 接收数据，并将数据写入到一个或多个 Channel 中，这些 Channel 是提前为 Source 配置过的。这也是 Source 最基本的职责。

本章我们将讨论 Flume 中封装的多种 Source 的设计和工作原理，以及如何才能最优化配置；我们也将介绍如何去编写自定义的 Source。

Source 的生命周期

Source 被命名为像任何其他部件一样通过配置文件进行配置的组件。Flume 的配置系统验证每个 Source 的配置和屏蔽错误配置的 Source。通过配置系统完成的验证相当少，不过，Flume 配置系统可以确保：

- 每个 Source 至少有一个正确配置的 Channel 连接它。
- 每个 Source 有一个定义的 type 参数。
- Source 是在 Agent 中 Sources 里面的活跃列表。

配置系统一旦验证通过一个 Source，就会实例化并且由 ConfigurationProvider 进行配置。如果 Source 缺少配置或者缺少需要的参数，这个 Source 将被屏蔽。Source 一旦成功配置，Flume 的生命周期管理系统将会尝试启动 Source。只有 Agent 自身停止或被杀死、或者 Agent 被用户重新配置，Source 才会停止。