

之间的相互通信”一节中已经讨论过, RPC Sink 被设计用来给 RPC Source 发送事件——Thrift Sink 发送给 Thrift Source, Avro Sink 发送给 Avro Source。RPC Sink 和 Source 是高扩展性的, 因为 Source 能接收大量的 Sink 或者 RPC 客户端发出的数据。尽管每个 RPC Sink 只能发送数据给一个 RPC Source, 但是每个 Agent 可以配置使用 Sink 组和 Sink 处理器, 来发送数据给多个其他的 Agent; 详情见第 6 章的“Sink 组和 Sink 处理器”。

在这部分中, 我们将要讨论 Flume 的 RPC Source, 包括配置的各个方面以及如何部署它们。

## Avro Source

Flume 主要的 RPC Source 是 *Avro Source*。Avro Source 被设计为高扩展的 RPC 服务器端, 能从其他的 Flume Agent 的 Avro Sink 或者使用 Flume 的 SDK 发送数据的客户端应用, 接收数据到一个 Flume Agent 中。Avro Source 和 Avro Sink 的组合代表了 Flume 内部的通信原理 (在 Flume Agent 之间)。Avro Source 的可扩展性结合 Channel 担当了缓冲器的角色, 使得 Flume Agent 能够处理重要的负载峰值。

Flume 的 Avro Source 使用 Netty-Avro inter-process 的通信 (IPC) 协议来通信。因此可以用 Java 或 JVM 语言发送数据到 Avro Source。如果你想从应用使用 Avro Source 发送事件给一个 Agent, 你可以利用 Flume SDK (详情见第 7 章的“Flume 客户端 SDK”一节) 或者嵌入式 Agent (详情见第 7 章的“嵌入式 Agent”节) 来完成。

Avro Source 可以配置用来从配置好输出压缩事件的 Avro Sink 中接收压缩的事件; 它也可以配置来确保接收任何客户端或 Sink 发送的使用 SSL 加密的数据。Avro Source 的配置参数在表 3-3 中将详细列出。

37 表3-3 Avro Source配置

配置参数	默认值	描述
<b>type</b>	-	Avro Source 的别名是 avro, 也可以使用完整类别名称, org.apache.flume.source.AvroSource (大小写敏感)
<b>bind</b>	-	绑定的 IP 地址或主机名。使用 0.0.0.0 绑定机器所有的接口
<b>port</b>	-	绑定的端口
<b>threads</b>	infinity	接收从客户端或 Avro Sink 传入的数据的最大工作线程的数量
<b>ssl</b>	false	SSL 应该启用吗? 如果设置为 true, 所有连接到 Source 的客户端都需要使用 SSL。如果启用了 SSL, keystore 和 keystore-password 参数是必需的