

## RPC Sink

正如前面章节解释的，集群拓扑需要 Flume Agent 发送数据到其他 Flume Agent。如果从一个 Flume Agent 发送数据到另一个 Agent，那么要使用 RPC Sink。RPC Sink 和对应的 RPC Source 使用相同的 RPC 协议。在第 3 章“Sink-to-Source 通信”一节中有关于 RPC Source 的更多细节。因此，RPC Sink 可以发送数据到 RPC Source。这种方法可以用于将数据从一个 Flume Agent 发送到另一个 Agent。因为 RPC Source 作为服务器监听指定的端口，那么可以实现几个 Flume Agent 发送数据到一个或多个使用对应的 RPC Sink 的 Flume Agent。

Flume 支持两种 RPC 系统，就像在第 3 章“Sink-to-Source 通信”一节解释的：Avro 和 Thrift [thrift\_ch5]。Avro 被认为是 Flume 主要的 RPC 格式，所以我们在这一节中将花更多的时间讨论 Avro Sink。我们还将简要查看 Thrift Sink，但因为 Avro Sink 更加成熟，它仍然是 Flume Agent 之间通信的推荐方法。两个 RPC Sink 共享一些基本配置参数，但是 Avro Sink 有更多的功能，因此有更多的配置选项。值得注意的是，Avro Sink 只能发送数据给 Flume Avro Source（或是建立在 Avro 的 Netty RPC 格式上的 Java 服务器），而 Thrift Sink 可以发送数据到 Flume Thrift Source 和任何语言编写的服务器进程，进程有 Thrift 服务器监听发送数据的 Sink 端口。

## Avro Sink

Avro Sink 使用 Avro 的 Netty-based RPC 协议发送数据到 Avro Source。它被实现为一个感知事务的 Flume NettyAvroRpcClient 的包装器。因此，它们共享一些配置参数。Avro Sink 可以批量发送事件到 Avro Source。这很重要，因为 Avro Source 不控制每个事务写出事件的数量——它将写出整个批次的事件。对于 File Channel，重要的是要以合理的批量写出，因为 Channel 将同步每个事务的数据文件到磁盘。为了避免太多昂贵的同步操作，批量大小必须适度大。

大型批量大小的另一个原因是，发送的每个批处理作为 RPC 调用。小批量可能意味着 RPC 开销在有效负载占据了很大比例，这应该避免。最好的批处理大小因为每个部署会有所不同。与 HDFS Sink 类似，它将取决于所使用的硬件、网络，甚至配置（如果在持有 Sink 和持有写入 Source 的两个 Agent 上批量大小都是大的，File Channel 将表现得更好）。

表 5-7 列出了 Avro Sink 的配置参数。

122