

Flume Agent 之间的相互通信

在下面几节中我们将看到，几乎总是需要一个 Flume Agent 发送数据到另一个 Agent。为了实现这一目标，Flume 内置了专门的 RPC sink-source 对。对于 agent-to-agent 通信，首选的 RPC sink-RPC source 对是 Avro Sink-Avro Source 对。

要从其他 Flume Agent 或客户端接收数据，Agent 接收数据可以被配置为使用 Avro Source，且发送数据的 Agent 必须配置用来运行 Avro Sink。Avro Sink 是发送事件到 Avro Source 的特定的 Sink。除了 Avro Sink，Flume RPC 客户端也可以发送事件到 Avro Source。Avro 通过 Avro RPC 协议，从其他 Flume Agent 或运行 Flume RPC 客户端的应用程序接收数据。单个 Avro Source 可以从大量的客户端或 Flume Agent 接收数据。即使单个 Avro Sink 只可以发送数据到单独一个 Avro Source，但是使用我们将在第 6 章“Sink 组和 Sink 处理器”一节讨论的 Sink 组和 Sink 处理器，可以从一个 Agent 发送数据到其他多个 Agent。

我们将在第 3 章“Sink-to-Source 通信”一节更详细地讨论这个问题。现在，重要的是要理解，可以从一个 Flume Agent 将数据发送到另一个 Agent，且通过客户端 API 也可以从自定义应用程序将数据发送到 Flume Agent（我们将在第 7 章进行讨论）。

复杂的流

如前所述，Flume Agent 可以有几个 Source、Sink 和 Channel，但是单个 Agent 中，这些组件的数量必须谨慎配置，以确保硬件不会崩溃。由于每个 Source 实际上可以写入到多个 Channel，事件可以很容易被复制，以确保每个事件到达多个目的地。然后，Sink 可以从 Channel 中删除数据，将数据推送到不同的目的地。

流是一个或多个的一系列 Agent，用来推送数据到另一个 Agent 并最终推送到存储或索引系统。在现实中，流可以任意地复杂，每三个组件作为多个流的一部分，且流包括多个 source-channel-sink 三件套。

◀ 18

一般来说，Flume 是为了从大量的服务器把数据发送到一个单独的 HDFS 集群。Flume 内置的 Avro 和 Thrift sink-source 对，可以用于从一个 Flume Agent 将数据发送到另一个。这允许用户设计从大量产生数据的应用程序的 fan-in-style 流。重要的是，要限制写数据到任意存储系统的应用程序的数量，以确保带大量数据写入的存储系统规模，并且可以处理写操作的突发性。