

数错误，RPC Source 都不会启动，例如绑定不到端口号，当 Source 尝试启动的时候会抛出异常。因为 Flume 的配置系统将每隔几秒重启组件，所以如果它的错误不再存在，配置成功的情况下，Source 将启动——例如，其他绑定到该端口的进程被杀或者释放端口。

更麻烦的部分是关于实际接收数据和往 Channel 写事件的代码。如果 Source 配置用来写入的其中一个 Channel 由于写满，会抛出 `ChannelException`，或这次事务量太大，Source 将会给客户端或 Sink 返回一个失败的状态，用来回调它并期望它重试。因为 RPC Source 通过线程池中的线程接收数据，线程异常只能导致线程终止。

在所有这些情况下，失败的真正提示只在这些记录异常的日志文件中。有时这些异常可能指明了一个大问题，例如在程序运行过程中耗尽资源（正如 `OutOfMemoryError`）。因此，监控 Flume Agent 生成的日志是很重要的，它能保证进程平稳运行。频繁抛出 `ChannelException` 意味着 Channel 分配远低于写入的速率，或者 Sink 清理 Channel 中数据的速度不够快。如果数量不足的 Sink 读取数据，可以增加 Sink 的数量，但是如果最终目的地本身不能处理该负载，则需要重新考虑能力。不管在什么样的情况下，错误可能导致程序重复执行但不会真的造成数据丢失，因为只有当数据真正成功写入下一阶段，事件才会从管道中移除。

43

HTTP Source

Flume 自带的 HTTP Source 可以通过 HTTP POST 接收事件。对于应用程序环境，它可能不能部署 Flume SDK 及它的依赖项，或客户端代码倾向于通过 HTTP 而不是 Flume 的 RPC 发送数据的情况，HTTP Source 可以用来将数据接收到 Flume 中。从客户端的角度来看，HTTP Source 表现得像 web 服务器一样能接收 Flume 事件。

如表 3-5 中展示的，HTTP Source 需要配置一些参数。这类 Source 的配置极其简单，还允许用户配置嵌入式的处理程序。

表3-5 HTTP Source配置

配置参数	默认值	描述
type	-	HTTP Source 的别名是 <code>http</code> ，也可以使用 FQCN， <code>org.apache.flume.source.HttpSource</code> （大小写敏感）
bind	-	绑定的 IP 地址 / 主机名。使用 <code>0.0.0.0</code> 绑定机器所有的接口
port	-	绑定的端口号