

表5-8 Thrift Sink配置

参数	默认值	描述
type	-	Thrift Sink 的别名是 thrift，也可以使用 FQCN <code>org.apache.flume.sink.ThriftSink</code>
hostname	-	托管 Agent 机器的主机名，Sink 应该连接的 Source 在该 Agent 上
port	-	Sink 连接的 Source 的监听端口
batch-size	100	每个 RPC 调用发送的事件数量。这也是在单次事务中 Sink 从 Channel 中读取的事件数量
request-timeout	20000	整个 RPC 调用成功完成的超时，以毫秒为单位
reset-connection-interval	-	Sink 从 Source 断开连接并重新连接之后的间隔
maxConnections	5	Sink 开放给 Source 连接的最大数量

除了 `maxConnections` 其余的所有参数都和 Avro Sink 中同名的参数有相同的意义。`maxConnections` 参数控制了每个 Sink 应该开放给 Source 连接的最大数量。如果需要，该参数可以用来调整该 Sink 使用的资源。第 3 章中“Thrift Source”一节展示了 Thrift Sink 配置的一个例子。

Morphline Solr Sink

Morphlines 是一个高度可扩展的 ETL（提取、转换和加载）框架，作为 Kite SDK[morphlines_ch5]的一部分发布。命令操作独立的记录。命令对输入记录做一个特定的转换，生成零或多个输出记录。几个命令可以连接在一起，且使一个命令的输出是下一个的输入，来执行数据的重量级转换。这样一组命令被称为 *Morphline*。Morphline Solr Sink 整合 Morphline 框架和提供 Sink 的 Flume，可以从 Flume 管道传输数据到这样一个 Morphline（命令的组合）。最重要的 Morphline Solr Sink 的使用，是关于 morphlines 与 Apache Solr[solr] 搜索的集成。

Morphline Solr Sink 通过移动事件主体到记录的 `_attachment_body` 字段，转换每个 Flume 事件为一条记录。报头被填充到字段，命名和报头键一样，值为 Flume 事件设置的对应值。每一批的事件从 Channel 被读取，作为单个 morphline 事务的一部分被处理，所以如果一个事件在 morphline 任何地方失败，那么整个事务被回滚。

正如前面所提到的，Morphline Solr Sink 最重要的作用是，实际加载 Flume 事件到 Solr 索引。Morphline 库提供了一个称为 `loadSolr` 的命令，可以加载记录到 Solr。使用该