

添加更多的 Morphline Sink 可以增加加载到 Solr 的事件数量，直到 Solr 中可用容量用完。一旦读取事件的总数量达到了批量大小或时间，因为批量启动后时间超过 `batchDurationMillis` 参数的值（或者，像其他所有的 Sink、Channel 中没有事件了），Sink 提交完 Sink 事务接着提交 morphline 事务。任何失败将会导致 morphline 事务和 Channel 有关的事务回滚。

Sink 可以配置为使用配置文件，类似于例 5-8 展示的例子。

例5-8 Morphline Solr Sink配置的一个例子

```
agent.sinks = morphline
agent.channels = solrChannel

agent.sinks.morphline.type = morphlinesolr
agent.sinks.morphline.morphlineFile = /etc/morphline/conf/morphline.conf
agent.sinks.morphline.morphlineId = usingFlumeMorphline
agent.sinks.morphline.batchSize = 1000
agent.sinks.morphline.batchDurationMillis = 2500
agent.sinks.morphline.channel = solrChannel

agent.channels.solrChannel.type = memory
agent.channels.solrChannel.capacity = 100000
```

这个配置文件描述了一个 Agent，该 Agent 通过 Avro Source 从其他 Flume Agent 或应用程序接收数据，然后通过两个 Memory Channel 复制数据到 HDFS 和 Solr。这是一个设置的典型例子，用来保证使用 Flume 实时地搜索索引。

事件被复制到 HDFS 是很常见的，Source 将相同事件写到 Channel 以供给该 Sink 和 HDFS Sink 也是常见的。为了合并 Solr 分片和用 Solr 调解 HDFS 上的数据，HDFS 上的数据可以用作 MapReduce 作业的输入 [solr-on-hdfs]。

在这个 Agent 中，morphline sink 使用 morphline 配置文件 `/etc/morphline/conf/morphline.conf`，且使用 `FlumeMorphline` 读取 morphline 的 ID。Sink 使用 1000 的批量大小，但是

◀ 129

如果花费多于 2.5 秒的时间达到这个数量，Sink 就提交该批量操作。

应该使用第 8 章“部署自定义代码”一节展示的 plug-in 部署系统，将 morphline JARs 和 morphline 配置文件添加到 Flume 的环境变量中。

Elastic Search Sink

Flume 也可以使用 Elastic Search Sink 实时加载数据到 Elastic Search 来创建索引。Elastic Search Sink 需要用户安装 Elastic Search 客户端 JAR 的正确版本，且在环境变量中安装