

Sink 加载 Flume 事件到 Solr，使用的 morphline 配置文件必须将 loadSolr 列为最后一个命令。每个 Flume 事件经过每个配置命令的转换，并最终通过 loadSolr 的命令被加载到 Solr。在加载可搜索的文本到 Solr 之前，转换可以用来解析文本为复杂格式表示的数据，例如 PDF、XML 等。在我们进入详细谈论 Solr 的用例之前，让我们看看表 5-9 中 Morphline Solr Sink 的配置参数。

表5-9 Morphline Solr Sink配置

配置	默认值	描述
type	-	Morphline Solr Sink 的别名是 morphlinesolr，也可以使用 FQCN org.apache.flume.sink.solr.morphline.MorphlineSolrSink
morphlineFile	-	从 morphline 读取文件的完整路径
morphlineId	-	如果在相同文件中有多个 morphline，Sink 应该使用来处理事件的 morphline 的 ID
batchSize	1000	每次批处理写入事件的数量
batchDurationMillis	1000	批处理被认为完成（如果有 loadSolr 命令，且加载到 Solr）之后的毫秒值时间

Morphline Solr Sink 的别名是 morphlinesolr，FQCN 是 org.apache.flume.sink.solr.morphline.MorphlineSolrSink，这两个都可以使用。morphlineFile 参数指定了 Sink 使用的 morphline 读取文件的路径。

Morphlines 框架允许 morphline 文件包含多个 morphline，每个用一个 ID 表示。如果文件中不止有一个 morphline，morphlineId 参数指定了文件使用的 morphline 的 ID。

batchSize 参数指定了一个事务中，Sink 从 Channel 读取事件的最大数量（在 Channel 中读取完最后一个事件，如果 Channel 中有更少的可用事件，批处理被认为立即完成）。这也是一个 morphline 会话中事件的总数量。在固定的持续时间之后也可以关闭 morphline 会话。该持续时间为 batchDurationMillis 参数指定的毫秒值。

为了在 Solr 上索引数据，每个插入到 Solr 的记录需要一个唯一的键来识别记录。对于 Solr 集合 [solr-unique-key]，作为键的字段使用 schema.xml 文件中 <uniqueKey> 标签所指定的值。在这个例子中，我们假设指定的唯一键的字段是 id。我们首先看一下用来加载数据到 Solr 的 morphline 文件（在例 5-7 中展示）。