

为具有相同行为的 HBase Sink 写一个序列化器是可能的。例 5-6 展示了需要实现的接口。

例5-6 HBase Sink的序列化器接口

```
package org.apache.flume.sink.hbase;

public interface HbaseEventSerializer extends Configurable,
    ConfigurableComponent {
    public void initialize(Event event, byte[] columnFamily);
    public List<Row> getActions();
    public List<Increment> getIncrements();
    public void close();
}
```

HBase Sink 的序列化器没有全局的初始化方法；HBase Put 和 Increment 不需要关于表的信息，这是由 Flume 直接处理的。相反，HbaseEventSerializer 中的 initialize 方法，和 AsyncHbaseEventSerializer 的 setEvent 方法做同样的事情，传递下一个将被序列化的事件和需要序列化的列族。getActions 和 getIncrements 方法和 AsyncHbaseEventSerializer 中同名的方法做同样的事情，唯一的区别是使用的 API。initialize 方法被调用传给事件后，Sink 立即调用这些方法。close 方法可以做任何需要清理的工作。只有当 Sink 即将停止才调用这个方法。

这两类序列化器都是 flume-ng-hbase-sink 包的一部分，你可以按照如下所示在 *pom.xml* 的依赖中添加它。

```
<dependency>
  <groupId>org.apache.flume.flume-ng-sinks</groupId>
  <artifactId>flume-ng-hbase-sink</artifactId>
  <version>1.5.0</version>
</dependency>
```

121



HBase 版本

HBase 0.96 之前，Apache 没发布针对 Hadoop 2 编译的 HBase JARs。Apache Flume 中的 *hadoop-2* 配置文件默认针对 HBase 0.94.2 编译。不幸的是，这个针对 Hadoop 2 建立的 HBase 版本在 Maven Central 上是不可用的。所以，用户必须使用 Hadoop 2 配置文件构建 HBase，以确保 Flume 针对 HBase 的正确版本建立。默认的配置针对 HBase 0.92.1 构建。Flume 也可以使用 *hbase-98* 配置文件针对 HBase 0.98.x 和 Hadoop 2.4.0 构建。如果使用任何其他版本的 HBase，必须要确保 HBase 的正确版本在环境变量中，Flume *pom.xml* 文件被修改为指向 HBase 的正确版本。