

编写预编码数据的一个很好的例子是 Flume 自带的 `AvroEventSerializer` 的工作方式。如果传入的事件是 Avro 格式，可以使用 Avro API 将它编码为一个字节数组，且设置值为事件主体。可以使用这种序列化器写入 HDFS，只是简单地写数据，从而使 HDFS 上 Avro container 文件中的原始 Avro 格式的数据可用。

HBase Sink

HBase 累积实时数据已越来越受欢迎，Flume 支持写入到 HBase。Flume 有两类 HBase Sink，*HBase Sink* 和 *Async HBase Sink*，它们的实现略有不同，但配置非常相似。HBase Sink 使用 HBase 客户端 API 来写数据到 HBase。因此，HBase Sink 更容易与 HBase 有线协议变化去同步的。

HBase 客户端 API 是阻塞的，所以 HBase Sink 逐个向 HBase 集群发送事件。另一方面，Async HBase Sink 使用 `AsyncHBase API[asynchbase]`，该 API 是非阻塞的且使用多线程写数据到 HBase。所以，大多数情况下，Async HBase Sink 有可能给出更好的性能。但是 HBase Sink 支持安全 HBase，这个是 Async HBase Sink 所不支持的。这两类 Sink 都支持序列化器，允许用户使用自定义逻辑转换 Flume 事件为 HBase 友好型对象。在这一节，我们将讨论 HBase sink 的两种类型，以及对于这两类 Sink 如何实现序列化器。

连接到一个或多个 HBase 集群的两类 HBase sink，在 Flume 配置文件或环境变量中的第一个 *hbase-site.xml* 文件中指明了它们的 Quorum。比起 HDFS Sink，HBase Sink 配置参数更少。两类 Sink 公共的参数展示在表 5-5 中。

表5-5 两类HBase sink的配置

参数	默认值	描述
type	-	HBase Sink 的别名是 <code>hbase</code> ，也可以使用 FQCN <code>org.apache.flume.sink.hbase.HBaseSink</code> 。Async HBase Sink 的别名是 <code>asynchbase</code> ，也可以使用 FQCN <code>org.apache.flume.sink.hbase.AsyncHBaseSink</code>
table	-	Sink 写入事件的表。该表必须在 HBase 中已存在——Flume 不会创建表
columnFamily	-	创建的列所在的列族。Flume 也不会创建列族，列族在 HBase 中必须存在