

参数	默认值	描述
batchSize	100	每次批处理写入事件的数量
zookeeperQuorum	-	HBase 集群使用的 quorum 中 Zookeeper 服务器的列表
znodeParent	/hbase	Zookeeper quorum 上 HBase 集群使用的父节点
serializer	H B a s e S i n k 的 使用的序列化器的 FQCN。在下一节讨论 SimpleHbaseEventSerializer, 如何写序列化器的更多细节 A s y n c H B a s e S i n k 的 SimpleAsyncHbaseEventSerializer	
serializer.*	-	传递给序列化器的配置参数的列表

115

这两类 Sink 接受这些参数，且关于这些参数有相同的特性。像我们至今见到多的所有其他组件，Sink 的 type 参数可以是别名或者 Sink 的实际 FQCN：HBase Sink 是 hbase 或 org.apache.flume.sink.hbase.HBaseSink，Async HBase Sink 是 asynchbase 或 org.apache.flume.sink.hbase.AsyncHBaseSink。这两类 HBase sinks 只能写入 table 和 columnFamily 参数所指定的一张表和一个列族。

这两类 HBase sink 都是批量写事件。batchSize 参数控制了每个事务写出事件的最大数量（在 Channel 中提供了足够可用的事件，如果 Channel 中有更少的事件，批处理会被认为立即完成）。每次批处理提交 Channel 的每个事务。当且仅当该批次中所有的事件被成功写入 HBase，该事务才被提交。为批处理的大小确定一个数值之前，尝试不同的值是很重要的，因为最合适的值取决于传入的确切数据、HBase 集群拓扑、模式设计、表是如何分裂为不同的 Region 服务器和网络体系结构的。在 HBase 文档或 *HBase: The Definitive Guide* [hbase-book] 中可以查找到 HBase 集群和 HBase 模式的设计细节。

默认情况下，两类 Sink 将在环境变量中查找客户端配置文件 *hbase-site.xml*，并且使用其中的信息连接 HBase。这带来了一个有趣的问题——Agent 中的每个 Sink 看到的是相同的环境变量，这意味着它不可能写数据到多个 HBase 集群。FLume 允许用户通过配置覆盖环境变量。

Sink 接受 zookeeperQuorum 参数，该参数接受逗号分隔的主机名和端口列表。主机名和端口指定为下面的格式：hostname1:port1、hostname2:port1（所有的服务器必须使用相同端口，这是 HBase 要求的）。Sink 也接受另外一个参数 znodeParent，为 HBase 集群使用的父节点（这通常不需要改变）。

116