

101

表5-4 HDFS Sink配置

参数	默认值	描述
type	-	HDFS Sink 的别名是 <code>hdfs</code> 。也可以使用 FQCN <code>org.apache.flume.sink.hdfs.HDFSEventSink</code>
hdfs.path		Sink 应该写入的目录路径。允许转义序列
<code>hdfs.filePrefix</code>	FlumeData	文件名的前缀。文件名包含的这个前缀在数字和名字之后
<code>hdfs.fileSuffix</code>	-	文件名使用的后缀。如果需要扩展，则应该明确指定 “.”
<code>hdfs.inUsePrefix</code>	-	HDFS Sink 正在写入的文件使用的文件名前缀
<code>hdfs.inUseSuffix</code>	.tmp	HDFS Sink 正在写入的文件使用的文件名后缀
<code>hdfs.timeZone</code>	-	创建 bucket 路径使用的时区
<code>hdfs.rollInterval</code>	30	文件保存之前的秒值时间。设置为 0 禁用基于时间的文件保存
<code>hdfs.rollSize</code>	1024	在文件保存之前写入文件的事件的最大值。设置为 0 禁用基于事件计算的文件保存
<code>hdfs.batchSize</code>	100	每批次写入事件的最大值
<code>hdfs.idleTimeout</code>	0	连续事件到一个未关闭的文件要等待的最大时间周期的秒值。设置为 0 禁用该项
<code>hdfs.fileType</code>	SequenceFile	使用的文件格式。可以是 <code>SequenceFile</code> 、 <code>DataStream</code> 或 <code>CompressedStream</code> 的一种
<code>hdfs.codec</code>	-	用来压缩文件的压缩编码。可以是 <code>gzip</code> 、 <code>bzip2</code> 、 <code>lzo</code> 、 <code>lzop</code> 或 <code>snappy</code> 中的一种
<code>hdfs.maxOpenFiles</code>	5000	HDFS Sink 一次可以保持打开文件的最大数量
<code>hdfs.callTimeout</code>	10000	每个 HDFS 操作超时前等待的毫秒延时
<code>hdfs.threadsPoolSize</code>	10	线程池中执行 HDFS 操作的线程数量
<div>102</div> <code>hdfs.rollTimerPoolSize</code>	1	基于 <code>hdfs.rollInterval</code> 和 <code>hdfs.idleTimeout</code> 参数的线程池中保存 HDFS 文件的线程数量
<code>hdfs.kerberosPrincipal</code>	-	登录到 Kerberos key 分布中心 (KDC) (使用的安全 HDFS) 的所用的 Kerberos 主体