

参数	默认值	描述
<code>hdfs.kerberosKeytab</code>	-	使用 <code>hdfs.kerberosPrincipal</code> 登录到 KDC（使用的安全 HDFS）的 keytab 文件的路径
<code>hdfs.proxyUser</code>	-	Flume 应该模仿的用户，如果需要。如果设置为 <code>none</code> ，Flume 以当前用户写入数据
<code>hdfs.useLocalTimeStamp</code>	<code>false</code>	如果设置为 <code>true</code> ，HDFS Sink 将使用现在 Agent 的时间戳做基于时间的分桶
<code>hdfs.round</code>	<code>false</code>	表明事件的时间戳是否应向下取整
<code>hdfs.roundUnit</code>	<code>second</code>	<code>hdfs.roundValue</code> 配置参数的单位（可以是秒、分钟或小时）
<code>hdfs.roundValue</code>	<code>1</code>	<code>hdfs.roundUnit</code> 参数指定的时间戳将舍去的多个参数的单位的最大值
<code>serializer</code>	<code>TEXT</code>	使用的序列化器。其他内置序列化器可以通过设置该值为 <code>AVRO_EVENT</code> 或 <code>HEADER_AND_TEXT</code> 来使用。为了使用自定义类，该值应该为序列化器 <code>Builder</code> 类的完整类别名称， <code>Builder</code> 类实现了 <code>EventSerializer\$Builder</code> 。序列化器本身必须实现 <code>EventSerializer</code> 。见下一节介绍
<code>serializer.*</code>	-	传递给序列化器的配置参数。见下一节介绍

为了使用 HDFS Sink，设置 Sink 的 `type` 参数的值为 `hdfs` 或使用 FQCN `org.apache.flume.sink.hdfs.HDFSEventSink`。

HDFS Sink 需要用户提供一个文件写入的目录路径。目录路径应该设置为 `hdfs.path` 参数的值。该参数支持分桶，在第 4 章“理解 bucket”一节中解释过。因此，该路径实际上不需要指定一个确切的目录，而是基于指定的分桶使用一组目录。HDFS Sink 创建的每个文件的名称，是根据 HDFS 配置中一系列参数指定的特定模式来的。

`hdfs.filePrefix` 参数指定了文件名使用的前缀。在该参数中分桶是允许的，但是只有当第一个文件创建，基于时间的分桶才会完成，并对 `bucket` 中所有的文件保持相同，所以最好要避免使用基于时间的分桶，并且在文件前缀中使用基于报头的分桶。

103

`hdfs.fileSuffix` 参数可以用来指明文件名的后缀。通常，通过使用这个参数指定文件的扩展。Flume 将不会自动在扩展之前添加句点 (.)，所以如果需要例如“txt”的扩展名，那么该参数的值需要被指定为 `.txt`。最终，文件名将会由文件前缀、文件后缀和中间的