

当 HDFS Sink 从 Channel 读取一个事件，它读取主题报头的值，并利用配置参数，将路径的转义序列替换为指定主题报头的值。例如，一个主题报头的值为 `inputData` 的事件，将写出到 HDFS 路径为 `/Data/Flume/inputData` 目录中的文件。另一个主题报头的值为 `logData` 的事件，将写出到 HDFS 路径为 `/Data/Flume/logData` 目录中的文件。该配置参数经常用来替换主机拦截器插入的主机信息，或者更经常的是，用来替换时间戳拦截器或生成事件的系统插入的时间戳。

HDFS Sink 提供了很强大的时间戳的转义。它支持几个使用时间戳事件报头值的转义序列。为了使用时间戳用于转义，Sink 对于报头为 `timestamp` 的键，预期为一个数字的值（或配置中 `hdfs.useLocalTimestamp` 参数必须设置为 `true`）。HDFS Sink 将时间戳的值认为是 Epoch，定义为自 1970 年 1 月 1 日以来 00:00:00 UTC 格式总的秒数，不包括闰秒。基于该报头的值，Sink 可以替换多个转义序列为相应的值。基于时间戳的转义序列在表 5-2 中列出。

表5-2 HDFS Sink时间戳转义序列

转义序列	替换值
<code>%t</code>	毫秒值的 Epoch 时间
<code>%s</code>	秒值的 Epoch 时间
<code>%H</code>	24 小时时钟指定的一天当中的小时值 (00..23)
<code>%I</code>	12 小时时钟指定的小时值 (01..12)
<code>%M</code>	每小时的分钟数 (00..59)
<code>%S</code>	每分钟的秒数 (00..59)
<code>%k</code>	和 <code>%H</code> 一样，但是最开始移除一个 0 (0..23)
<code>%p</code>	AM 或 PM（或者当前区域它们的相等值）
<code>%z</code>	时区指定为自 UTC 以来用 <code>+hhmm</code> 格式表示的差值（例如， <code>+0530</code> 表示印度标准时间，比 GMT 时间提前 5 个小时 30 分钟）
<code>%a</code>	当前区域的周的星期简称（例如 Mon、Tue 等）
<code>%A</code>	当前区域的周的星期全称（例如 Monday、Tuesday 等）
<code>%b</code>	当前区域的月的简称（例如 Jan、Feb 等）
<code>%B</code>	当前区域的月的全称（例如 January、February 等）
<code>%d</code>	月的天数（例如 01、02 等）