

表5-3 HDFS Sink round-down参数

参数	默认值	描述
<code>hdfs.round</code>	<code>False</code>	表明事件的时间戳是否应向下取整
<code>hdfs.roundValue</code>	<code>1</code>	时间戳将取整这个参数多个值中的最大值，以 <code>hdfs.roundUnit</code> 参数指定的值为单位
<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）

如果 `hdfs.round` 参数设置为 `true`，那么 HDFS Sink 从事件报头中读取时间戳，并将它向下取整到 `byhdfs.roundValue` 指定的多个值中最高的一个，以 `hdfs.roundUnit` 指定的值为单位。当使用了 `%t` 转义序列，向下取整的值就不能使用。让我们看看这里展示的 Sink 的配置参数：

```
agent.sinks.hdfsSink.hdfs.path = /Data/Flume/%{topic}/%Y/%m/%d/%H/%M
agent.sinks.hdfsSink.hdfs.round = true
agent.sinks.hdfsSink.hdfs.roundUnit = minute
agent.sinks.hdfsSink.hdfs.roundValue = 10
```

如果一个事件的时间戳是 2:35 PM on September 1, 2014，HDFS Sink 向下取整的时间戳为 2:30 PM on September 1, 2014，并将这个事件写出到 `/Data/Flume/UsingFlume/2014/09/01/14/30` 目录。实际上，位于 2:30:00:00:000 PM 和 2:39:59:59:999 之间的事件都是写出到相同的目录。使用类似的配置原理（就像下面展示的），用户可以每两小时就有 bucket。

```
agent.sinks.hdfsSink.hdfs.path = /Data/Flume/%{topic}/%Y/%m/%d/%H/%M
agent.sinks.hdfsSink.hdfs.round = true
agent.sinks.hdfsSink.hdfs.roundUnit = hour
agent.sinks.hdfsSink.hdfs.roundValue = 2
```

配置 HDFS Sink

HDFS Sink 有很多配置参数，允许用户在一个很好的粒度级别控制 Sink 的表现。表 5-4 列出了配置 HDFS Sink 使用的参数。