

转义序列	替换值
%C	当前区域指定的天，日期和时间（例如，Sun Feb 9 14:05:45 2014）
%m	月的数值表示（例如，January 将会是 01）
%D	以 mm/dd/yy 表示的日期
%y	世纪开始以来的年（00..99）
%Y	当前年份（例如 2013、2014 等）

多个转义序列可以在相同的路径中使用。这实质上意味着用户可以根据下面展示的日期、小时、分钟等，将它们的消息分桶（bucket）：

```
agent.sinks.hdfsSink.hdfs.path = /Data/Flume/{topic}/%Y/%m/%d/%H
```

这个例子对于每一个主题，将把所有时间戳为相同小时的事件写到相同目录内。如果一个事件主题名为 UsingFlume，时间戳是 2:35 PM on September 1, 2014，它将写入到 */Data/Flume/UsingFlume/2014/09/01/14* 目录。

通常基于时间的分桶，在该小时后的几分钟，通常使用 Oozie 或 cron job（在 Flume 管道中引发延时）运行加载数据到系统中的查询，用来加载数据到像 Hive 或 Impala 的系统中。这种分桶的一个重要方面是，有更老时间戳的数据有可能被写入到 bucket，因为 Sink 使用的时间戳是基于事件的报头——在 Flume Agent 中，如果时间戳在事件创建时或通过一个拦截器被插入，Agent 故障或网络延时会造成事件到达 HDFS Sink 的时间比时间戳晚很多。如果一个 Hive 查询是基于时间分桶加载数据，一定会造成写入的延时。如果只有最新的目录被扫描，那么目录中更老时间戳的新数据可能从不会被加载。

HDFS Sink 有一个配置参数，`hdfs.useLocalTimestamp`，如果设置为 `true`，那么强制 Sink 使用运行 Agent 机器的系统时间；它可以用来确保数据基于当地时间戳写入。这个参数需要谨慎使用，因为 HDFS 集群和 Flume Agent 可能不同步并且可能不会有精确的相同时间。

这里描述的基于时间分桶的配置有一个问题，可以按照每年、天、小时、分钟或秒分桶，但是每几分钟创建一个新 bucket 是不可能的；例如，每十分钟一个新 bucket。如果用户想每十分钟进行分桶，那么他需要每分钟有一个 bucket，意味着他将以 HDFS 上众多数量的小文件而收场。但是 HDFS Sink 确实提供了一种通过配置参数的组合来实现的方法。所需的三个配置参数列在表 5-3 中。