

Agent 失败和数据丢失

如果一个 Agent 失败，或存储系统不可达，会发生什么？当一个 Agent 或最终目的地变得不可达，Agent 会在 Sink 看到错误的位置写入数据时结束。Sink 会看到连接失败，或下一阶段或存储系统的 ACK 丢失（或存储系统的客户端 API 抛出异常）。当 Sink 遇到异常，表明数据可能没有被写入，Sink 利用 Channel 回滚事务。因为传入的数据流并没有停止，这导致 Channel 大小增加，最终填满 Channel。当这个 Agent 的 Channel 被填满，Avro Source（或任何其他写数据到该 Channel 的 Source）开始在 Channel 的 Put 操作上得到 `ChannelExceptions` 异常，通过返回一个错误导致 Source 回退到前一阶段。进而导致前一阶段的 Sink 回滚得到事务，导致那个阶段的 Channel 大小增长。这个场景如图 2-10 所示。

26

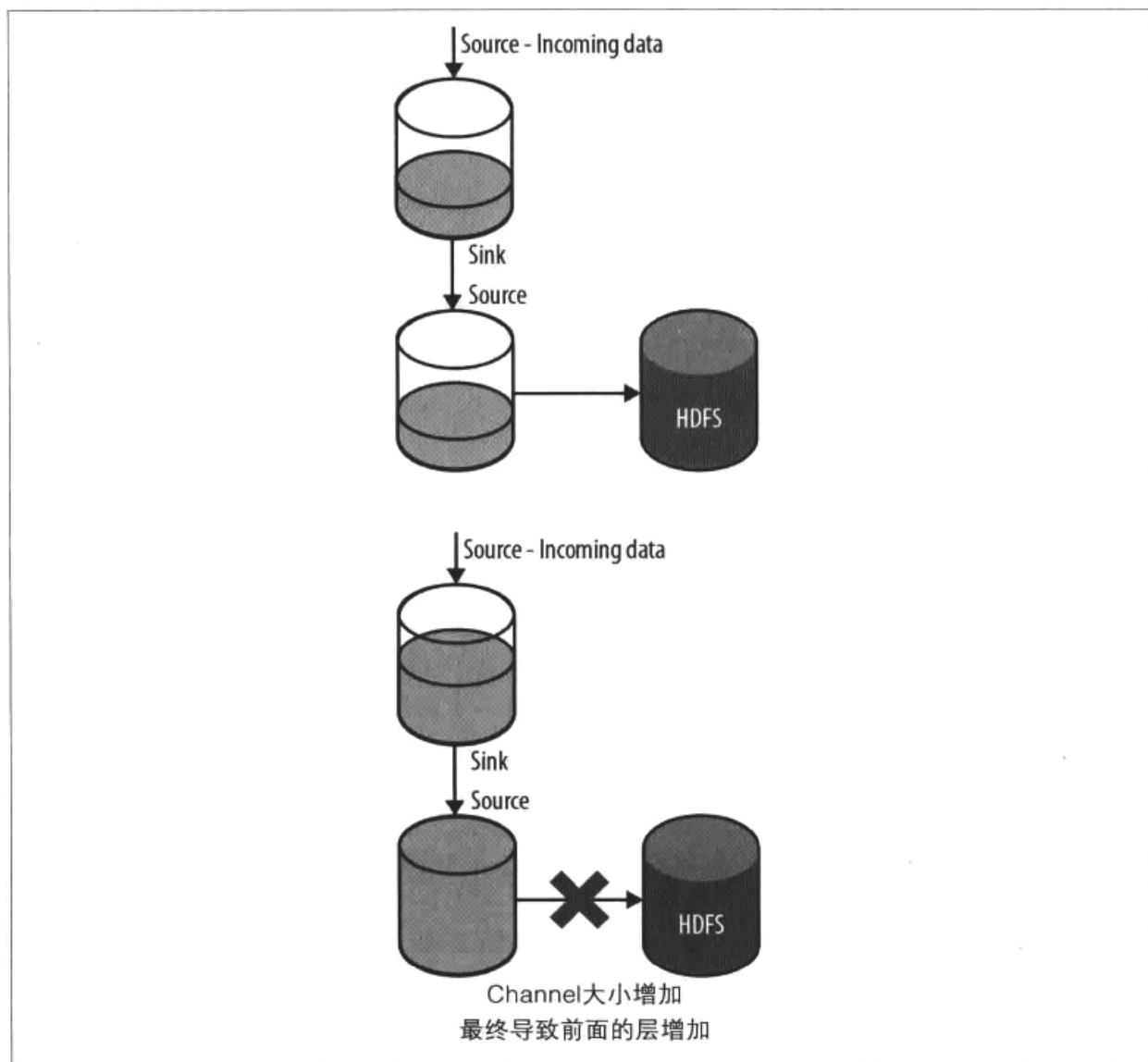


图2-10 失败导致Channel被填满