Avro Sink 和 Flume RPC 客户端可以配置用于在发送数据给 Avro Source 之前压缩数据。如果数据在广域网或数据中心之间传输时,这是非常有用的,能够减少使用的带宽。目前,Avro Source 仅对 RPC 支持 zlib 压缩。为了使 Avro Source 能以压缩的形式接收数据,需要将 compression-type 参数设置为 deflate。如果这个参数没有设置或者设置为 none,Flume 将不会解压数据;这可能导致事件积压在前一阶段,因为 Source 将不能解析压缩数据,会给前一阶段返回错误,将导致该阶段一直重试。

39



Avro Source 和压缩

如果 compression-type 参数设置为 deflate,那么传入的数据必须被压缩,否则 Source 不能解析传入的数据。Sink 或 Flume 客户端发送数据时,必须配置为压缩将要发送的数据。因此,如果压缩和未压缩的数据被传输到相同的 Flume Agent, agent 应该能运行两种 Avro Source,一种用来接收压缩的数据,另一种用来接收未压缩的数据。

还有需要注意的是,当 Avro Sink 或 Flume RPC 客户端发送数据到 Avro Source 时,数据是分批被压缩的,而不是按照每个事件,因此这可能会提供更好的压缩率,减少压缩时内存的使用情况。

下面是一个配置了 SSL 和压缩的 Avro Source 的例子。如果要禁用 SSL, 移除 ssl 参数即可,剩下的 SSL 相关的参数将会被忽略:

agent.sources = avroSrc
agent.channels = memChannel

agent.sources.avroSrc.type = avro
agent.sources.avroSrc.channels = memChannel

Bind to all interfaces// 绑定到所有的接口 agent.sources.avroSrc.bind = 0.0.0.0

agent.sources.avroSrc.port = 4353

Removing the next line will disable SSL

// 删除下一行将禁用 SSL

agent.sources.avroSrc.ssl = true

agent.sources.avroSrc.keystore = /tmp/keystore.jks

agent.sources.avroSrc.keystore-password = UsingFlume

agent.sources.avroSrc.keystore-type = jks