



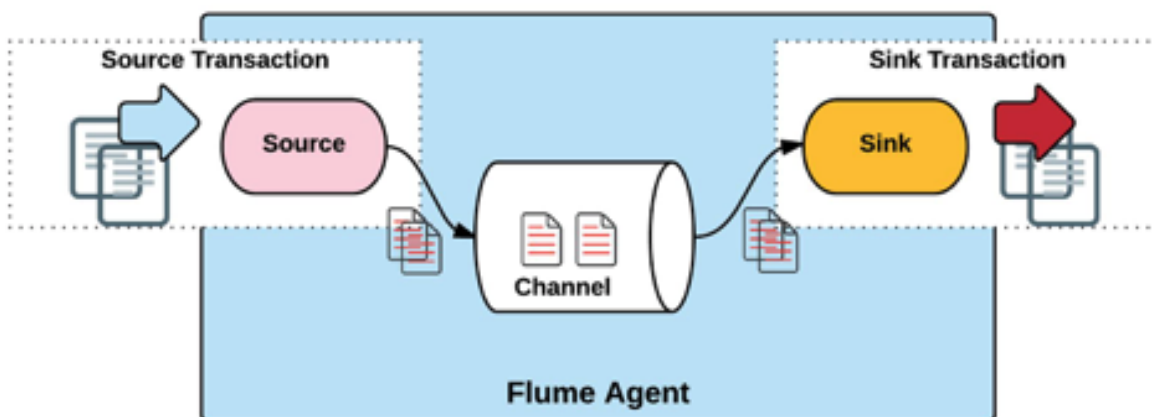
Data Science Academy

www.datascienceacademy.com.br

Data Lake – Design, Projeto e Integração

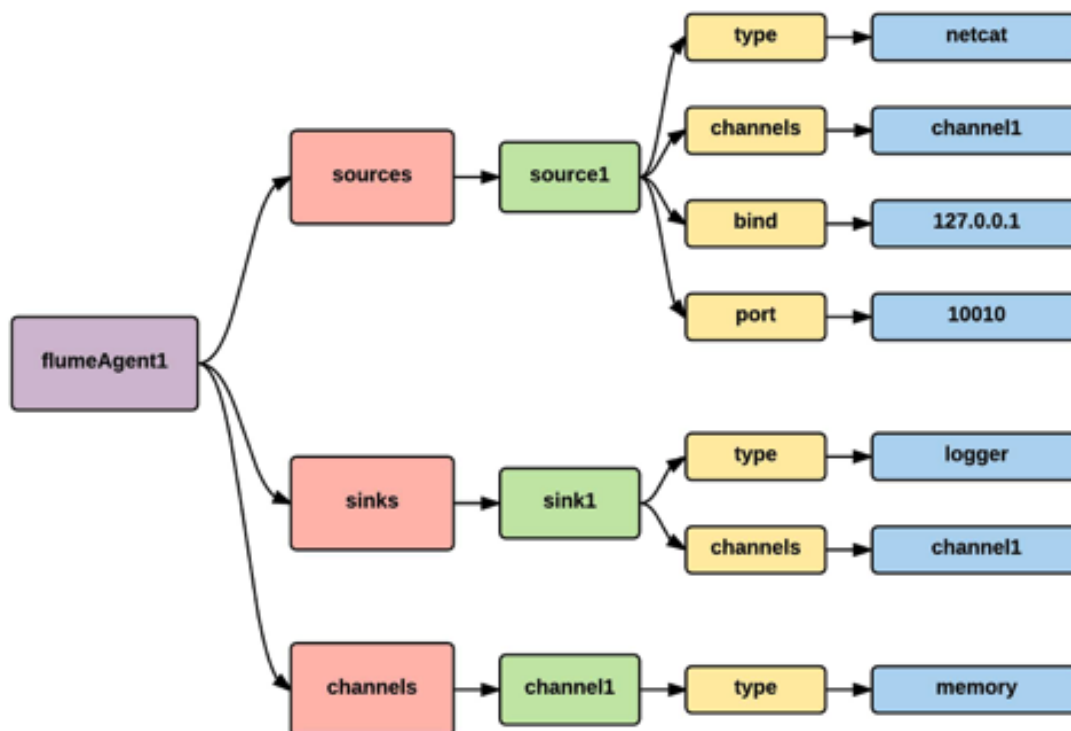
Gerenciamento de Transações dos Eventos e Configuração Flume

Ao longo das aulas anteriores, estudamos, de fato, aspectos de transação em vários estágios. A figura a seguir resume essas discussões de uma maneira mais clara:



Essa figura mostra que os dados de entrada de um cliente ou coletor iniciam a transação do agente e isso é denominado como **Source Tx** (Source Transaction) na figura. O **Source Tx** termina logo após o evento persistir no canal e o reconhecimento (aceite) recebido. Uma segunda transação é denominada como **Sink Tx**, que começa com os dados sendo pesquisados pelo coletor e quando os dados são transferidos com sucesso, o canal usa a confirmação para remover os dados do canal. O Flume possui gerenciamento de transações em todos os aspectos e, de acordo com o caso de uso, vários níveis de confiabilidade podem ser definidos no canal que decide como o comportamento da transação (Sink Tx) é realizado.

O Flume pode ser totalmente configurado usando arquivos de configuração (a exemplo do Hadoop). Uma única imagem fala mais de mil palavras, então gostaríamos de explicar a configuração do Flume usando a figura a seguir.



Abaixo um exemplo de configuração do Flume, considerando a figura acima:



Componentes Flume

flumeAgent1.sources = source1

flumeAgent1.channels = channel1

flumeAgent1.sinks = sink1

Definindo e Configurando Source 1

flumeAgent1.sources.source1.type = netcat

flumeAgent1.sources.source1.channels = channel1

flumeAgent1.sources.source1.bind = 127.0.0.1

flumeAgent1.sources.source1.port = 10010

Definindo e Configurando Sink 1

flumeAgent1.sinks.sink1.type = logger

flumeAgent1.sinks.sink1.channels = channel1

Definindo e Configurando Channel 1

flumeAgent1.channels.channel1.type = memory

Outros Agentes Flume

flumeAgent2.sources = source2

Referências:

<https://flume.apache.org/FlumeDeveloperGuide.html>