



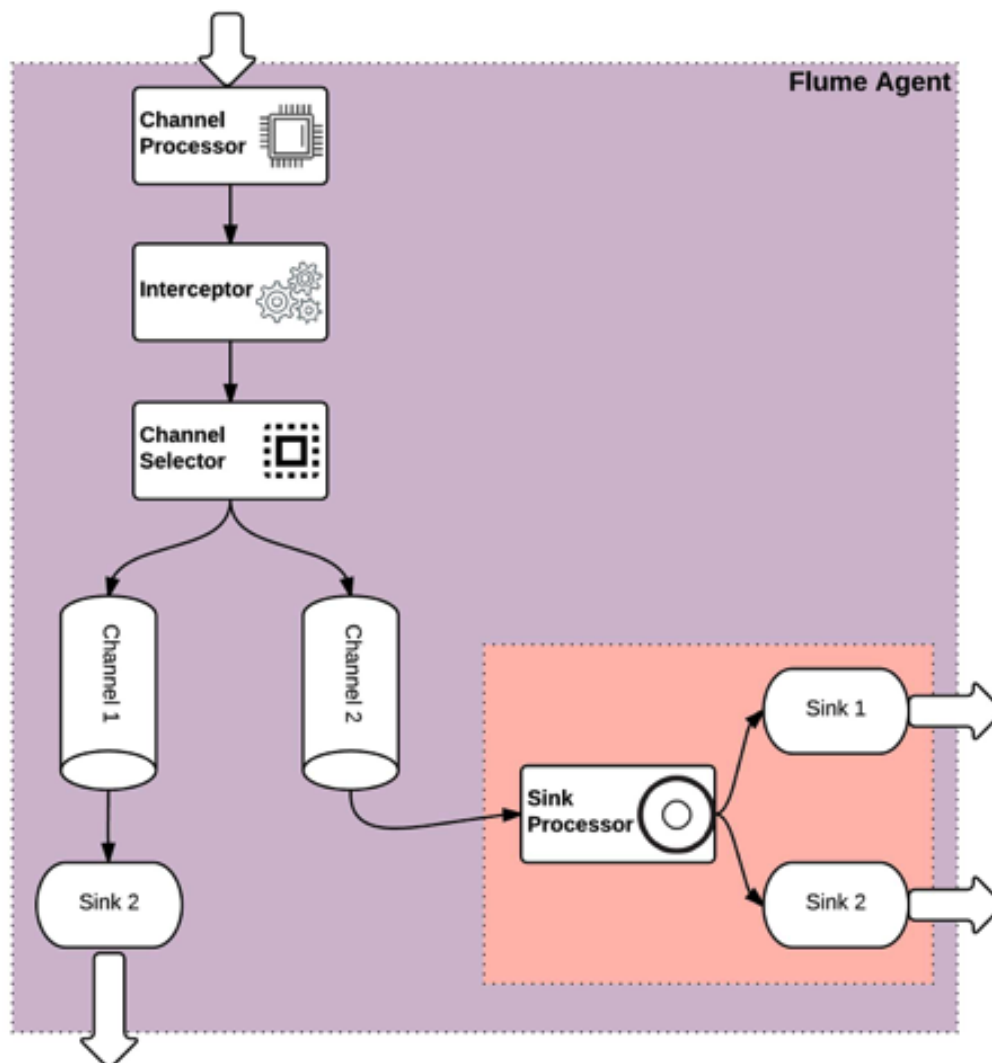
# Data Science Academy

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

## Data Lake – Design, Projeto e Integração

### Outros Componentes do Processamento de Streaming

Além dos componentes principais do Flume, existem outros componentes adicionais de grande importância. A figura a seguir mostra todos esses componentes trabalhando juntos:



Vamos compreender como esses componentes ajudarão você a projetar o arranjo correto dos componentes Flume para executar seu caso de uso e resolver seu problema de negócio.

## Processador de Canal (Channel Processor)

Começamos pelo Processador de Canal (Channel Processor). Conforme pode ser visto, a fonte envia os eventos para o processador do canal. Toda fonte tem seu próprio processador de canal e para persistir o evento no canal, a fonte delega o trabalho para o processador de canal, que realmente realiza a tarefa de persistir de acordo com o tipo de canal.

## Interceptor

Outro importante componente é o Interceptor. Como visto na figura anterior, o processador de canal passa os eventos para o interceptor. O seletor de canais tem uma instância de seu Interceptor anexado a ele. Os interceptores agem sobre o evento logo após o evento ser gerado e antes de enviá-lo ao canal. O Flume tem a capacidade de modificar e filtrar eventos com a ajuda desses interceptores. Interceptores são principalmente classificados em:

- **Interceptors Built-in** - Como o nome diz esses interceptors estão embutidos no Flume. Alguns exemplos são:
  - **Timestamp Interceptor:** Insere o cabeçalho de tempo em milissegundos quando o evento é acionado.
  - **Host Interceptor:** Insere o host ou IP da máquina na qual o agente está em execução.
  - **UUID Interceptor:** Insere um identificador universal exclusivo para cada evento.
  - **Pesquisar e substituir o interceptor:** com base em expressões regulares, ele fornece a funcionalidade de pesquisa e substituição de funcionalidade.
- **Interceptors Personalizados:** Ao olhar para o evento, ele pode inserir cabeçalhos ou transformar o evento. É necessário escrever um código Java simples para um interceptor personalizado.

## Seletor de Canais (Channel Selector)

Como mostrado na figura anterior, o seletor de canais ajuda a selecionar o canal correto com base nos critérios definidos. Muito semelhante aos interceptors, o seletor de canais é classificado em duas categorias principais: Built-in Channel Selector e Custom Channel Selector.

## Grupos de Coletores (Sink Groups)

A confiabilidade é um dos principais princípios de arquitetura em que o Flume foi desenvolvido. Para cuidar das falhas, os componentes do canal podem ser organizados de várias maneiras. Em caso de falhas durante a transferência de dados do canal para o coletor, o coletor pode ser configurado de forma balanceada ou com failover de carga. Para alcançar este pré-requisito precisamos configurar um Sink Group. Um Sink Group, como o próprio nome sugere, é um agrupamento lógico de sinks (coletores). O grupo definido ou denominado participa de um balanceamento de carga ou de failover.

Alguns dos aspectos de um Grupo de Coletores que são dignos de nota são os seguintes: Um coletor só pode ser associado a um Grupo de Coletores de cada vez. Todos os coletores têm seu próprio grupo de coletores padrão ao qual pertencem. Se não especificado, cada coletor pertence a esse grupo de coletores padrão e se comporta como uma passagem para eventos. O Grupo de Coletores pode ser desativado a qualquer momento, mas a desativação não afeta os coletores que participam do Grupo. Na configuração Flume, o Sink Group é um elemento de nível superior.

## Processador de Coletores (Sink Processor)

O Processador de Coletores determina como o Grupo de Coletores funcionará e atingirá os cenários de balanceamento de carga ou de failover requeridos pela garantia de confiabilidade acordada para sua configuração do Flume. O Processador de Coletores também é um componente de nível superior na configuração Flume. O Processador de Coletores é classificado em dois tipos: Built-in Sink Processor e Custom Sink Processor.



## Context Routing

Como explicado anteriormente, o evento possui duas partes principais, a saber, Header e Payload. Os valores de cabeçalho (par chave/valor) podem ser usados e, conseqüentemente, o roteamento definido. Dois componentes onde a seleção de roteamento pode ser decidida são:

**Canal:** Um canal pode ser selecionado de acordo com os valores do cabeçalho.

**Coletor:** Os valores de cabeçalho podem ser usados para tomar decisões e selecionar o coletor correto. Além disso, dentro do coletor, diferentes operações podem ser executadas escrevendo-se um coletor personalizado que pode fazer o que seu caso de uso exigir. Existem alguns valores de cabeçalho padrão que também podem ser usados para fazer coisas sofisticadas para o seu caso de uso.

Ufa! Depois de todo esse conteúdo, hora de ver o Flume em ação e você vai perceber que tudo que leu nas últimas aulas, é mais importante do que você pensa!

Referências:

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