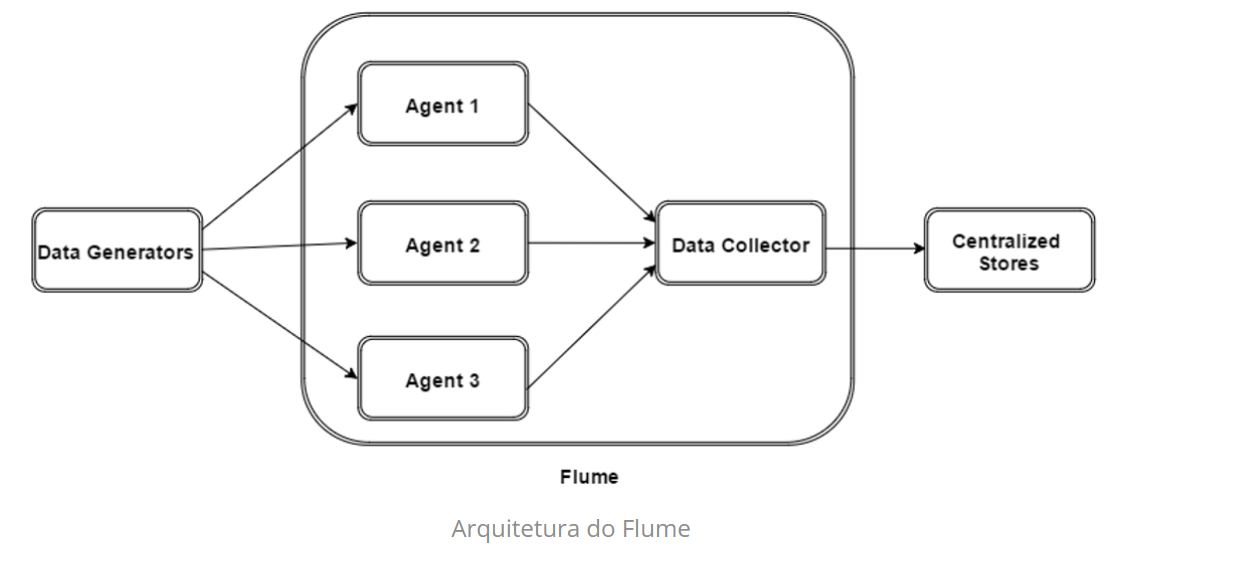
Carga de dados para o pipeline

Flume é uma ferramenta do apache hadoop que coleta uma grande quantidade de dados e os envia para o HDFS (Hadoop Distributed File System), sendo geralmente usado para dados não estruturados. Um agente Flume roda em uma Java Virtual Machine e possui 3 componentes:

* Source: responsável pela entrada de dados;
* Channel: armazena os dados que passam do source para o sink (parecido com uma fila de espera).
* Sink: responsável por enviar os dados ao destino/ saída.



https://www.codigofluente.com.br/wp-content/uploads/2018/05/Flume-Arquitetura.png

Os pontos de interesse no Flume são a sua capacidade de transportar dados não relacionais para o HDFS (Hadoop Distributed File System) de forma simples e automatizada. Lembrando que seu uso não se limita apenas ao HDFS, sendo possível enviar também dados para um arquivo ou banco de dados e entre outros. O Apache Flume é uma ferramenta flexível que tem um sistema distribuído, confiável e disponível, tornando-a uma ferramenta simples e poderosa.

https://www.codigofluente.com.br/ingestao-de-dados-com-o-flume/

<https://imasters.com.br/desenvolvimento/introducao-ao-apache-flume>

<https://www.dezyre.com/article/sqoop-vs-flume-battle-of-the-hadoop-etl-tools-/176>