**Apache Sqoop**

É uma das ferramentas que possibilitam a transferência entre os bancos de dados relacionais (SQL), bancos de dados não estruturados (NoSQL), MainFrames e datawarehouse (DW), isso tudo de forma automática, tornando-o um grande aliado na hora de fazer essa operação. É importante salientar que essa ferramenta consegue transferir quantidades de dados massivas e atua de forma bidirecional, sendo que tudo que entra é possível transformar, após o processo de importação dos dados, você pode exportá-los de forma em que já venham transformados pelo MapReduce.

Diagrama

Descrição gerada automaticamente

**Apache Flume**

O apache flume já nasceu como um sistema construído para armazenar e mover grandes quantidades de dados, seja qual for a fonte que deseja extrair, o flume será seu aliado nessa operação e como já citado, diferentemente do Sqoop que em geral apenas trabalha com banco de dados relacional, ele trabalha com dados não estruturados também como por exemplo, imagens, textos, áudio e muito mais. Sua arquitetura é simples, tem como base esses três elementos: Source, Channel e sink, um fato interessante é que é possível decidir como a ferramenta irá distribuir os dados, então aliado com um pouco de programação em Java, é possível decidir como vai querer distribuir os dados no sistema. Apesar de ter um funcionamento simples, isso não torna o sistema menos completo, com esse horizonte, é possível combinar diferentes fluxos e com isso é possível adaptar-se a quase qualquer cenário, com essa grande flexibilidade, a ferramenta se torna um grande aliado no cenário Big Data, onde é necessário adaptar-se ao cenário, ao processamento e volumetria de dados.

<https://bdm.unb.br/bitstream/10483/11335/1/2015_GuilhermedeLimaBernardes.pdf>

<https://pt.linkedin.com/pulse/uma-introdu%C3%A7%C3%A3o-ao-apache-flume-marco-antonio-pereira-mba>

https://www.ufpb.br/datascience/contents/paginas/fcd/engenharia-de-dados-com-hadoop-e-spark-versao-2.0