



Data Science Academy

www.datascienceacademy.com.br

Data Lake – Design, Projeto e Integração

Flume Vantagens e Desvantagens

Algumas das principais vantagens do Apache Flume, que fizeram com que essa tecnologia seja escolhida em projetos de coleta e armazenamento de dados em tempo real:

- Open-source.
- Uma documentação muito boa, com muitos exemplos e padrões de como estes podem ser aplicados.
- Alto rendimento com baixa latência.
- Configuração declarativa.
- Distribuído por padrão.
- Altamente confiável, disponível e escalonável (horizontalmente).
- Altamente extensível e personalizável.
- Instalação, operação e manutenção menos dispendiosas.
- Suporte incorporado para uma variedade de sistemas de origem e destino.
- Suporte a transações está embutido.
- Capacidade de obter dados de vários servidores e armazenar no HDFS.
- Suporta diferentes fluxos de dados, como multiple-hop, fan-out, fan-in entre outros.
- Bom suporte de integração com uma variedade de ferramentas de monitoramento existentes.

Mas o Flume também apresenta desvantagens:

- Não garante que o alcance da mensagem seja exclusivo (mensagens duplicadas podem aparecer em alguns momentos, em muitos cenários).
- Para uma empresa, dimensionar o hardware para uma implementação típica do Flume pode ser complicado e, na maioria dos casos, é tentativa e erro. Devido a isso, seu aspecto de escalabilidade é frequentemente questionado.
- A taxa de transferência que o Flume pode manipular depende do armazenamento de suporte do canal.
- Portanto, escalabilidade e confiabilidade estão em questão quando a escolha do data store não é feita de maneira inteligente, considerando todos os fatores.



Toda tecnologia possui vantagens e desvantagens, mas no caso do Flume, as vantagens são maiores.