



Data Science Academy

www.datascienceacademy.com.br

Formação Cientista de Dados

Projeto com Feedback 6

Processamento de Logs em Tempo Real
Com Flume, Spark Streaming e Hbase

Na computação, um arquivo de log é um arquivo que registra eventos que ocorrem em um sistema operacional ou outro software, ou mensagens entre diferentes usuários de um software de comunicação. Logging é o ato de manter um log. No caso mais simples, as mensagens são gravadas em um único arquivo de log. Exemplos de arquivos de log, incluem:

- Proxy log
- Transaction log
- Event log
- Message log
- Application log
- Web server log

A análise de logs tem várias aplicações práticas e os principais casos de uso de processamento de log de servidor web por exemplo, incluem:

- Monitoramento de integridade de aplicativos
- Fraude - Segurança
- Padrão do usuário (sessão de um fluxo de cliques)
- Experiência de usuário
- Triagem de Suporte
- Coleta de dados métricos

Neste projeto de Big Data, você deverá construir um pipeline de monitoramento em tempo real de logs de aplicação usando Apache Flume para aquisição de dados, Spark Streaming para processamento e Apache Hbase para armazenamento.

Os detalhes sobre nossas recomendações de ferramenta e datasets e como enviar seu projeto para análise, você encontra quando concluir a avaliação final.