



Análise de Sentimentos Twitter - Guerra da Ucrânia



Utilizando Hive e PySpark

Grupo 1

- Arthur de Castilho Nascimento 11270674
- Barbara Ellen S. Finardi 11207769
- João Matheus Pinheiro 11207939
- Kleber Araújo San Galo Curvelo 11207922

O objetivo deste trabalho foi utilizar ferramentas de Big Data que ajudam no armazenamento e processamento de grandes volumes de dados de qualquer estrutura.

O conjunto de dados analisado trata-se de tweets sobre a Guerra da Ucrânia tem tamanho de 10 GB. Foi utilizado o Apache Hive (sistema de data warehouse para Apache Hadoop) e a API para Python do Spark, o PySpark.



Passo 1

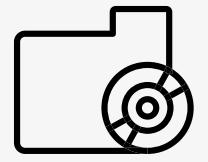


Captura de
Dados

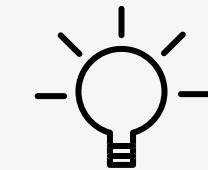


- Utilização da API do Kaggle para captura de tweets
- Injeção dos tweets no Hive





Hive permite o resumo de dados, consultas e análise de dados. Consultas de hive são escritas em HiveQL, que é uma linguagem de consulta semelhante ao SQL.



Ele permite que você projete estrutura em grandes volumes de dados sem estrutura. Depois de definir a estrutura, você pode usar o HiveQL para consultar os dados sem conhecimento de Java ou do MapReduce.





Passo 2

Análise de
Sentimentos



Utilizando um script Python fazemos a análise de sentimento das tuplas.



Passo 3

Captura de Resultados





Conforme as análises são processadas seus resultados são armazenados no banco.

Obrigado!

ACH2066 Tópicos Especiais em Bancos de Dados

