

CADERNO DE RESPOSTAS DA ATIVIDADE PRÁTICA DE:

Projetos BIG DATA

ALUNO: emanuel rosa zolet - RU: 4557826

**ANO**

**2024**

Caderno de Resposta Elaborado por:

Prof. MSc. Guilherme Ditzel Patriota

|  |  |
| --- | --- |
| **Prática 01 – Somatório de IDs** | |
| **Questão – Qual o valor da soma de todos os campos “id” dos filmes classificados como negativos para o banco de dados “imdb-reviews-pt-br.csv”** | |
| **ENUNCIADO: Nessa prática você deverá descobrir, utilizando sua máquina virtual com o Hadoop ou Spark ou PySpark, qual o valor da soma de todos os campos “id” dos filmes classificados como negativos para o banco de dados “imdb-reviews-pt-br.csv”.** | |
| 1. **Apresentação do Código (não esquecer do identificador pessoal):**)   **Figura 1: Captura de tela demonstrando o código completo do módulo python responsável por somar os ids dos filmes.** | 1. **Apresentação das Imagens/Print do resultado (não esquecer do identificador):**     Figura 2: Captura de tela com a saída de informação do console. |
| 1. **Responda à pergunta: Qual o valor da soma de todos os campos “id” dos filmes classificados como negativos?**   **Resposta: A soma total de todos dos os campos id negativos são 459568555** | |

|  |  |
| --- | --- |
| **Prática 02 – Diferença do número de palavras totais de português para inglês dos textos negativos** | |
| **Questão – Contar palavras dos textos negativos e achar diferença de quantidade.** | |
| **ENUNCIADO:** Nessa prática você deverá contar todas as palavras existentes nos textos negativos (Português e Inglês) e então deverá encontrar quantas palavras a mais, no total, os textos em português possuem.  Para tal, crie um script em Python ou Scala e rode-o com sua máquina virtual Hadoop ou Spark ou PySpark, como feito na prática 1.  É necessário se preocupar em filtrar corretamente as avaliações de filmes para que apenas os textos marcados como negativos sejam contabilizados. | |
| 1. **Apresentação do Código (não esquecer do identificador pessoal):**     **Figura 3: Captura de tela do módulo python, utilizado para a realização do trabalho.** | 1. **Apresentação das Imagens/Print do resultado (não esquecer do identificador):**     Figura 4: Captura de Tela, visualização dos dados via terminal |
| 1. **Responda à pergunta: Qual o número total de palavras a mais que os textos negativos em português possuem em comparação com a soma total das palavras dos textos negativos em inglês, independentemente de serem repetidas?**   **Resposta:** O número total de palavras a mais é de 54976 | |