

CADERNO DE RESPOSTAS DA ATIVIDADE PRÁTICA DE:

Projetos BIG DATA

ALUNO:   
MOACIR DOMINGOS DA sILVA jUNIOR   
ru 3539252

**ANO**

**2023**

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: Apresentação do Código** | 1. **Apresentação das Imagens/Print do resultado (não esquecer do identificador):**     Figura : Print do Resultado do código executado no  Google Colab |
| 1. **Responda à pergunta: Qual o valor da soma de todos os campos “id” dos filmes classificados como negativos?**   **Resposta:** **A soma de todos os campos ‘Id” dos filmes classificados como negativo é** **247015948.** | |

|  |  |
| --- | --- |
| **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: Apresentação do Código** | 1. **Apresentação das Imagens/Print do resultado (não esquecer do identificador):**     Figura : Print do Resultado do código executado no  Google Colab |
| 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:** **Os textos negativos em português possuem 160085 palavras a mais do que os textos negativos em inglês.** | |