

Hands- on com Map Reduce – Parte I

1. Considere como dataset o arquivo **movies.txt**, disponível na pasta data do MS Teams compartilhada com sua turma. Esse arquivo apresenta o seguinte layout (**user_id, movie_id, rating, timestamp**). Com base nisso, implemente o MapReduce em Python para fornecer uma lista de todos os filmes, ordenados de forma ascendente pela popularidade de cada um deles. Considere que para nós, a popularidade de um filme NÃO É medida pelo número de avaliações ou pelas notas recebidas. Assuma que a popularidade é dada pela média aritmética das avaliações que o filme recebeu. Com base nos seus resultados obtidos deve ser possível responder facilmente as seguintes perguntas:
 - * Qual o filme de maior popularidade?
 - * Qual o filme menos popular?
2. Considere como dataset o arquivo **Marvel--graph.txt**, disponível na pasta data do MS Teams compartilhada com sua turma. Esse arquivo apresenta em cada linha o ID de um dado herói, seguido dos IDs de seus amigos heróis. Com base nisso, implemente o MapReduce em Python para encontrar quem é o herói mais popular, isto é, nesse caso, quem deles **tem mais amigos heróis**.
3. Considere como dataset o arquivo **movies.txt**, disponível na pasta data do MS Teams compartilhada com sua turma. Esse arquivo apresenta o seguinte layout (**user_id, movie_id, rating, timestamp**). Com base nisso, implemente o MapReduce em Python para fornecer a lista dos DEZ filmes mais populares, **ordenados de forma decedente**.