

## Grupo 05

Nome (RA):

Lucas Alves Racoci (156331)

Luiz Fernando Rodrigues da Fonseca (156475)

Rafael Zane (157079)

Rodrigo Noronha Máximo (157209)

Tema

Análise da Issue UMBRELLA SPARK-9953: ML Vector, Matrix semantic equality + hashCode

Descrição da Issue:

Issue umbrella para trocar os tipos locais de álgebra linear da MLlib's (Vector, Matrix, DenseVector, SparseVector, DenseMatrix, SparseMatrix) para que implementem igualdade semântica e levem em consideração transposição, zeros e esparsidade. Para isso é necessário atualizar o hashCode.

Plano de Trabalho:

Estudar a issue para entender melhor qual a requisição descrita, analisar as issues relacionadas com essa e os comentários para entender quais os progressos realizados e dificuldades enfrentadas. Há uma discussão inicial sobre o tema no bug já resolvido SPARK-9750, que diz que a SparseMatrix deveria sobrescrever a igualdade para que duas instâncias da mesma matriz fossem iguais.

A UMBRELLA contém várias issues, algumas em progresso e outras fechadas ou resolvidas. A issue SPARK-9792 diz que a igualdade deveria ser definida semanticamente, não em termos de representação e está aberta. A issue SPARK-9919 diz que as matrizes deveriam respeitar as igualdades de Java e o hashCode, mas que seria preciso algumas modificações para tratar colisões e não calcular o hash com todos os valores, o que pode ser caro computacionalmente, e esta issue está em progresso. Há também duas issues que se referem a sobrecarregar o hash do Python, para seguir as mesmas convenções de Scala, mas que já foram fechadas ou resolvidas.