



**Título: G2P-DBSCAN: Estratégia de Particionamento de Dados e de Processamento Distribuído do DBSCAN com MapReduce**

**Data: 17/08/2015 Horário: 16h30min Local: Sala de Seminários Bloco 952**

Resumo: Clusterização é uma técnica de mineração de dados que agrupa elementos de um conjunto de dados de forma que os elementos que pertencem ao mesmo grupo são mais semelhantes entre si que entre elementos de outros grupos. Nesta dissertação nós estudamos o problema de processar o algoritmo de clusterização baseado em densidade DBSCAN de maneira distribuída através do paradigma MapReduce. Em processamentos distribuídos é importante que as partições de dados a serem processadas tenham tamanhos aproximadamente iguais, uma vez que o tempo total de processamento é delimitado pelo tempo que o nó com uma maior quantidade de dados leva para finalizar a computação dos dados a ele atribuídos. Por essa razão nós também propomos uma estratégia de particionamento de dados, chamada G2P, que busca distribuir o conjunto de dados de forma balanceada entre as partições e que leva em consideração as características do algoritmo DBSCAN. Mais especificamente, a estratégia G2P usa estruturas de grade e grafo para auxiliar na divisão do espaço em regiões de baixa densidade. Já o processamento distribuído do algoritmo DBSCAN se dá por meio de duas fases de processamento MapReduce e uma fase intermediária que identifica clusters que podem ter sido divididos em mais de uma partição,

chamados de candidatos à junção. A primeira fase de MapReduce aplica o algoritmo DSBCAN nas partições de dados individualmente, e a segunda verifica e corrige, caso necessário, os clusters candidatos à junção. Experimentos utilizando dados reais mostram que a estratégia G2P-DBSCAN se comporta melhor que a solução utilizada para comparação em todos os cenários considerados, tanto em tempo de execução quanto em qualidade das partições obtidas.

Banca:

- JAVAM DE CASTRO MACHADO (UFC - Orientador)
- JOSE ANTONIO FERNANDES DE MACÊDO (MDCC/UFC)
- JOSÉ NEUMAN DE SOUZA (MDCC/UFC)
- ALTIGRAN SOARES DA SILVA (UFAM)