

Proposta de Trabalho Final

Introdução ao Geoprocessamento (SER-300)

Vitor Conrado Faria Gomes

Orientadores: Dra. Karine Ferreira e Dr. Gilberto Ribeiro

Título: Avaliação de Desempenho de SGBD Matriciais com Séries Temporais de Imagens de Sensoriamento Remoto

Os avanços tecnológicos dos equipamentos de aquisição de dados científicos tem contribuído para uma maior disponibilidade de dados nas mais diversas áreas da Ciência. Essa maior disponibilidade de dados trás grandes desafios de armazenamento, gerenciamento e análise dos dados. Novas tecnologias de bancos de dados tem surgido para atender essas questões. Na área de Observação da Terra, uma das tecnologias que tem sido empregadas em projetos de pesquisa são os Sistemas Gerenciadores de Banco de Dados Matriciais (SGBD-M). Esses SGBDs são projetados para manipular dados densos e/ou esparsos, utilizando um modelo de dados mais apropriado aos dados científicos, que são as matrizes multidimensionais. Além disso, lidam de forma nativa com questões como vizinhança, particionamento dos dados (horizontal e vertical) e replicação, além de operarem em um ambiente de cluster computacional.

Neste contexto, a proposta deste trabalho é explorar os principais SGBD-M a fim de (i) identificar seus recursos para a manipulação e o armazenamento de dados matriciais; e (ii) avaliar o desempenho desses sistemas quando utilizados para o gerenciamento e análise de Séries Temporais de Imagens de Sensoriamento Remoto. Planeja-se trabalhar com os principais SGBD-M de código livre disponíveis atualmente: SciDB [1], RasDaMan [2] e TileDB [3]. Para a análise de desempenho, objetiva-se ir além da coleta do tempo necessário para a realização de tarefas, para isso, serão medidos: (i) tempo e uso de recursos do hardware (CPU, memória, rede e disco) para a realização de operações como carga, processamento e recuperação dos dados; (ii) espaço em disco utilizado para armazenar o conjunto de dados; (iii) capacidade de reduzir o tempo a partir do aumento de recursos de hardware disponíveis (escalabilidade). Para os testes, será utilizado um conjunto de imagens do sensor MODIS de 2000 a 2016 e um conjunto de operações suportadas pelos SGBD-M.

Referências

[1] BROWN, P. G. Overview of SciDB: Large Scale Array Storage, Processing and Analysis. Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data. 2010.

[2] BAUMANN, P. et al. The Multidimensional Database System RasDaMan. Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data. 1998.

[3] PAPADOPOULOS, S. et al. The TileDB Array Data Storage Manager. Proceedings of the VLDB Endowment. 2016.