

Software para Distribuição dos Processos TIMESAT - SDPT

Carlos Padonavi¹

Embrapa, Pantanal, MS

Diego Bulhões Moraes²

UFMS, Campo Grande, MS

Guilherme Ribeiro de Carvalho³

UFMS, Campo Grande, MS

Hernanes de Oliveira Almeida⁴

UFMS, Campo Grande, MS

Renato Porfirio Ishii⁵

UFMS, Campo Grande, MS

O software TIMESAT foi concebido para fins de suavização de séries temporais de imagens de satélite e extração de parâmetros de sazonalidade, ajuste de função de pixels e visualização de imagens de satélite. Além disso, permite que o usuário faça filtragem e retire ruídos de uma série de imagens sazonais, por exemplo, identificar regiões inundadas, sem ruídos gerados por nuvens e reflexão da luz da água. Essas atividades exigem elevada capacidade computacional. Uma alternativa do TIMESAT para acelerar o processamento dos dados é permitir a parcelarização o processamento dos dados o TIMESAT permite a paralelização, especificando o número de núcleos de processamento, mas dependendo da quantidade e volume de dados das imagens, o processamento pode demorar dias. A fim de reduzir o tempo de processamento, está em desenvolvimento um software em Python capaz de distribuir os processos do TIMESAT em *cluster*⁶, o Software para Distribuição dos Processos TIMESAT (SDPT).

O TIMESAT possui diversas aplicações para tarefas específicas, porém, durante o projeto foram utilizadas as aplicações *TSF_process.x64* e *TSF_merge.x64*. O *TSF_process.x64* é responsável por dividir o arquivo de configuração inicial (arquivo que contém a lista da série de imagens a serem processadas e parâmetros de processamento) e processar as imagens utilizando as configurações especificadas. O *TSF_merge.x64* realiza a junção dos arquivos de saída após o processamento das imagens.

O primeiro procedimento realizado pelo SDPT durante a distribuição dos processos do TIMESAT é a divisão do arquivo de configuração inicial usando o *TSF_process.x64*, onde é passado como parâmetro o arquivo de configuração inicial e o um número inteiro que representa a quantidade de partes que esse arquivo será dividido (também podemos interpretar esse número como a quantidade de CPUs que processará a lista de imagens). Após a divisão, ocorre a distribuição dos processos no *cluster* utilizando o protocolo Secure Shell (SSH)⁷, onde o SDPT cria *n* sessões com os nodes⁸, no qual possui uma quantidade de CPUs livres em um node específico, sendo que cada sessão é responsável por executar remotamente um único processo *TSF_process.x64*.

¹carlos.padovani@embrapa.br.

²diegobulhoesm@gmail.com.

³guilhermercavalho512@gmail.com.

⁴hernanes.almeida@gmail.com.

⁵renato@facom.ufms.br .

⁶Conjunto de máquinas conectadas com objetivo de aumentar o desempenho ou disponibilidade.

⁷Protocolo de rede que permite login remoto de forma segura

⁸Menor unidade que compõe um *cluster*

Concluído o processo, são gerados arquivos do tipo **.tts* e **.tpa* (que podem variar conforme os dados de entrada contidos no arquivo de configuração) que serão utilizados pelo *TSF_merge.x64* para realizar a junção dos resultados, também executado de forma distribuída no *cluster*. Por fim, terminada a junção são obtidos como saída três tipos de dados: (1) dados de sazonalidade, **_TS.tpa*; (2) funções ajustadas, **_fit.tts*; e (3) dado original, **_raw.tts*. Na figura 1 ilustra o diagrama mostrando os passos executados durante o processamento usando o SDPT, desde o recebimento do arquivo de configuração até o envio dos resultados obtidos pelo processamento das imagens.

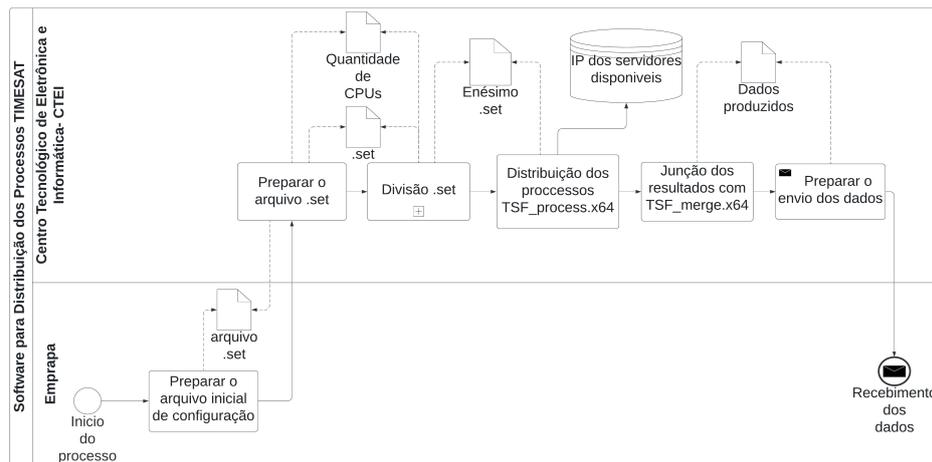


Figura 1: -Diagrama com os passos executados pelo SDPT, imagem criada via Lucidchart.

A medida que mais núcleos de processamento estão disponíveis no *cluster*, mais processos podem ser executados, possibilitando assim a obtenção dos resultados em um menor intervalo de tempo. Apesar da divisão de tarefas, ainda enfrentamos um gargalo na execução do processo, gerado principalmente pela junção dos dados, ao executar o *TSF_merge.x64*, que aguarda o processo de todos os dados serem finalizados para então realizar o *merge*. Como este programa não é distribuído e é executado apenas em uma máquina com poucos núcleos quando comparado a um *cluster*, o tempo de processamento é impactado. O que afeta diretamente o tempo final do todo processo.

A interação do usuário com o SDPT está sendo realizada via linha de comando, contudo está sendo desenvolvida uma aplicação web que possibilita pesquisador poderá agendar e analisar os resultados do processamento das imagens, Software para Distribuição dos Processos TIMESAT WEB (SDPTW). A ferramenta está sendo desenvolvida utilizando o NodeJS e invoca o SDPT, assim realizando as tarefas: gerar o arquivo de configuração inicial e monitoramento dos processos.

Referências

- [1] Coulouris, George F. Et Al. *Sistemas Distribuídos: Conceitos e Projeto*. 5. Ed. Porto Alegre, Rs: Bookman, 2013. Xvi, 1048P.
- [2] Eklundh, L., and Jönsson, P., 2017, *TIMESAT 3.3 with seasonal trend decomposition and parallel processing - Software Manual*. Lund University, 92 pp.
- [3] Eklundh, L., and Jönsson, P., 2012, *TIMESAT 3.2 with parallel processing - Software Manual*. Lund University, 88 pp..
- [4] Eklundh, L., and Jönsson, P., 2012, *TIMESAT 3.1 - Software Manual*. Lund University, 82 pp.