



Evento	Salão UFRGS 2020: SIC - XXXII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2020
Local	Virtual
Título	Investigando Tamanhos de Requisições de Entrada/Saída de Aplicações HPC em um Supercomputador
Autor	GESSICA FRANCIÉLE MENDONÇA AZEVEDO
Orientador	PHILIPPE OLIVIER ALEXANDRE NAVAUX

Investigando Tamanhos de Requisições de Entrada/Saída de Aplicações HPC em um Supercomputador

Gessica F. M. Azevedo, Philippe O. A. Navaux

Instituto de Informática - Universidade Federal do Rio Grande do Sul

Para avaliar novas técnicas de otimização para aplicações HPC (*High Performance Computing*), é necessário utilizar parâmetros próximos da realidade para que a validação, feita utilizando *benchmarks*, seja consistente. Nosso objetivo é identificar os tamanhos de requisições de Entrada e Saída (E/S) mais comuns para que os testes mantenham sua confiabilidade, já que estes parâmetros estarão de acordo com a realidade das aplicações HPC. Entre 2010 e 2013, o Argonne Leadership Computing Facility coletou dados de execução de uma variedade de aplicações científicas usando a ferramenta de caracterização de E/S chamada Darshan, no supercomputador Intrepid Blue Gene/P. Darshan intercepta o fluxo de funções de E/S e registra uma coleção de estatísticas para cada arquivo que é aberto. As informações coletadas permitem identificar tamanhos e padrões de acesso, operações e tempo gasto em operações de E/S. O foco neste trabalho foi em dados de 2012 gerado pela versão Darshan 2.0, resultando em 91.603 *jobs*. Estas informações foram agrupadas por mês e padrão de acesso, afim de facilitar a identificação do tamanho de requisição mais comum para cada padrão de acesso e seu comportamento. Ao longo do ano, foram observadas as estatísticas de mínimos, médias, máximas e quartis destes tamanhos de requisições e 22 padrões únicos de acesso. Retiramos, como amostra, apenas os seis padrões que mais ocorreram dentre os *jobs* das aplicações, para gerar uma visão geral dos tamanhos das requisições. Concluimos que destes seis padrões utilizados como amostra, cinco deles se mantêm com a tendência do tamanho médio das requisições entre 1KB e 1MB, estes padrões utilizam a interface POSIX. O único padrão que se destaca por sua alta variabilidade para os tamanhos médios foi o padrão que utiliza a interface MPI-IO, que disponibiliza otimizações como *data sieving* e *collective buffering*, o que pode justificar a diversidade destes tamanhos.