

Extração e Análise do Comportamento de I/O de Aplicações de Alto Desempenho

Valéria S. Girelli, Philippe O. A. Navaux

Instituto de informática - Universidade Federal do Rio Grande do Sul - Porto Alegre, Brasil

Introdução

Em sistemas de computação de alta performance (*High Performance Computing* - HPC), onde aplicações executam de forma paralela em centenas de nós de computação, os dados também podem ser acessados de forma paralela. Esse acesso pode ser feito de diversas formas, resultando em diferentes padrões de entrada e saída (*Input and Output* - I/O). Por exemplo, as requisições enviadas ao sistema de I/O podem ser muito pequenas, transferindo pequenas quantidades de dados que dificilmente compensam o custo de acessar os discos de dados que estão em um diferente conjunto de nós. Uma solução para este problema é agregar essas pequenas requisições, uma das possíveis otimizações que podem ser feitas no sistema de I/O. No entanto, para aplicar esse tipo de otimização precisa-se primeiramente compreender o comportamento de acesso a dados das aplicações.

Ferramentas como Darshan fornecem uma caracterização do comportamento de I/O das aplicações. O Darshan foi desenvolvido em Argonne Leadership Computing Facility (ALCF), e captura informações a respeito das operações de leitura e escrita a nível de aplicação. Portanto, com o objetivo de entender o comportamento global das operações de I/O em um supercomputador, este estudo analisou dados coletados utilizando o Darshan no supercomputador de ALCF, o Intrepid Blue Gene/P.

Metodologia

Durante os anos de 2010, 2012 e 2013, dados da execução de uma variedade de aplicações científicas foram coletados pelo Darshan no supercomputador Intrepid. O Darshan intercepta chamadas a funções de leitura e escrita e registra um conjunto de estatísticas para cada arquivo aberto pelas aplicações. As informações coletadas incluem padrões de acesso, tamanhos de acesso, contadores de operações e tempo gasto com operações de I/O. Neste estudo, foram analisadas as informações coletadas durante todo o ano de 2012 pelo Darshan, resultando em 91.994 execuções.

Análise e Resultados

Foi observado que 97,3% das operações de leitura e 99,2% das operações de escrita estavam sendo realizadas com POSIX, como mostra a Figura 1. Portanto, POSIX continua sendo mais amplamente utilizada se comparado com MPI-IO. Isso pode ser um problema, pois POSIX é uma interface que transfere ao programador toda a responsabilidade sobre a manutenção das operações de I/O. Interfaces de mais alto nível, como MPI-IO, auxiliam no acesso aos dados, assim como na aplicação de algumas otimizações.

A Figura 1 mostra também que o tamanho de acesso mais comum para operações de leitura foi entre 10KB e 100KB, representando 64,3% das requisições. O tamanho mais comum para operações de escrita foi de até 100 bytes, um tamanho consideravelmente pequeno e representado por 98,7% das requisições de escrita.

Para identificar se o comportamento observado era reflexo de apenas um conjunto de aplicações ou se ele representava o comportamento geral das aplicações que executam no Intrepid, analisamos então os tamanhos de acesso observado entre as 10 aplicações que fizeram a maior quantidade de operações de I/O.

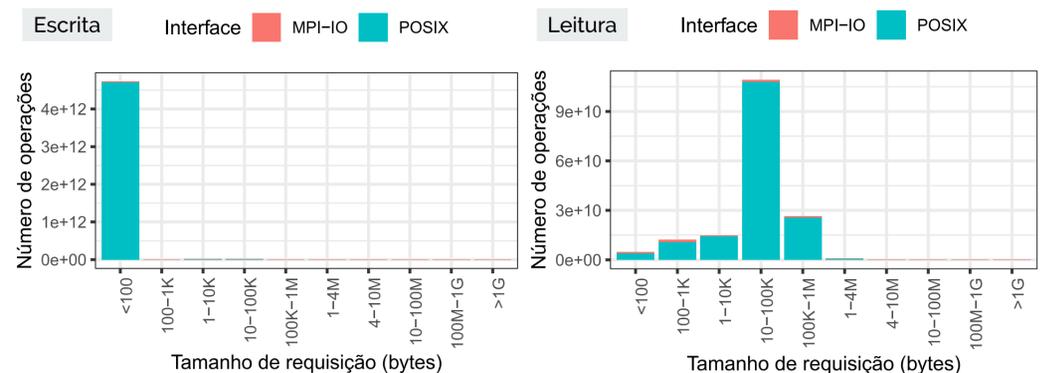


Figura 1: Distribuição do tamanho de acesso das requisições por interface. O eixo y tem diferentes escalas.

Observamos então que as 10 maiores responsáveis pelas operações de I/O pouco impactaram na quantidade de requisições feitas com MPI-IO. No entanto, a Figura 2 mostra que essas 10 aplicações foram responsáveis por 99,1% das requisições de escrita de até 100 bytes feitas com POSIX e por 99,5% das requisições de leitura feitas com POSIX entre 10KB e 100KB, os dois tamanhos de requisição mais observados.

Removendo essas 10 aplicações da análise e aplicando a mesma metodologia, o novo tamanho de acesso mais comum para operações de leitura passou a ser de 1KB a 10KB, representado por 31,5% das requisições. Dentre as requisições de escrita, requisições de até 100 bytes continuam sendo as mais observadas.

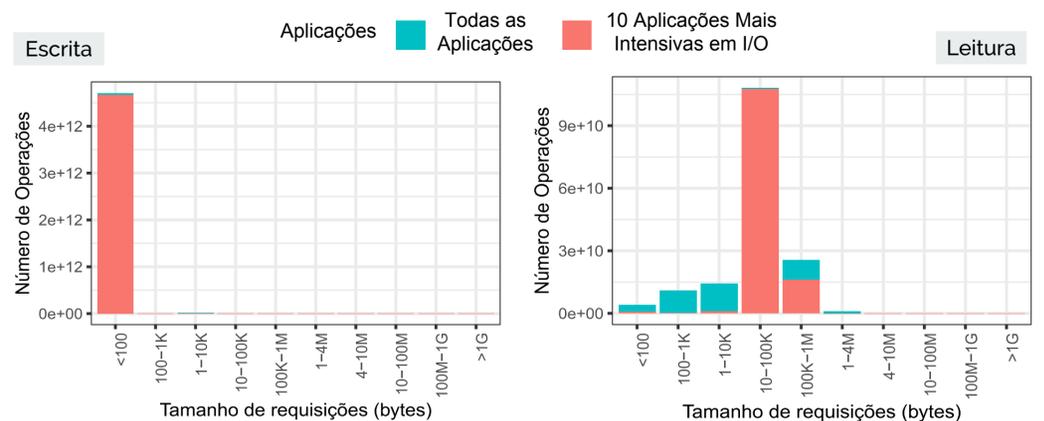


Figura 2: Proporção das requisições realizadas pelas 10 aplicações que mais fizeram operações de I/O. O eixo y tem diferentes escalas.

Conclusão

Analisando informações sobre a execução de aplicações durante um ano inteiro no supercomputador Intrepid, observamos que POSIX ainda é amplamente utilizada para realizar operações de I/O, superando 97% de utilização para operações de leitura e 99% para operações de escrita. Observa-se que a maior parte das aplicações não está tirando proveito de vantagens fornecidas por interfaces de mais alto nível, como MPI-IO.

Em geral, as requisições de escrita estão sendo realizadas em um tamanho consideravelmente pequeno. Essas requisições de até 100 bytes podem prejudicar significativamente o desempenho observado em operações de I/O, e isso normalmente está relacionado com a forma como a aplicação foi codificada.

Como trabalho futuro, iremos analisar a quantidade de dados transferida por cada tamanho de acesso observado, e identificar os tamanhos de acesso responsáveis pelas maiores transferências de dados. Além disso, queremos investigar o tempo de sistema gasto com operações feitas com cada tamanho.