LNCC
www.lncc.br
Laboratório Nacional de Computação Científica

Introdução E/S Paralela no SDUMONT
Darshan

Escola Supercomputador SDUMONT
Programa de Verão 2021

André Ramos Carneiro (andrerc@lncc.br)
Bruno Alves Fagundes (brunoaf@lncc.br)

Darshan

LNCC
www.lncc.br
Laboratório Nacional de Computação Científica

- Ferramenta escalável para caracterização das operações I/O
- Laboratório Nacional de Argonne
- Investigação ou tunning das operações de I/O
- Aplicações que utilizam mais de 25% do wallclok com operações de I/O merecem uma atenção especial.
- http://www.mcs.anl.gov/research/projects/darshan/

Darshan

Características:

- Baixo overhead
- MPI-IO e POSIX
- Mantém os dados coletados em um buffer de memória
- Não armazena informações completas
- Armazena os dados compactados
- Integração transparente com a maioria das aplicações

Darshan

LNCC
www.lncc.br
Laboratório Nacional de Computação Científica

darshan-runtime

Conjunto de bibliotecas e wrappers de compilação para interceptar e contabilizar as chamadas de sistema para acesso aos arquivos.

darshan-utils

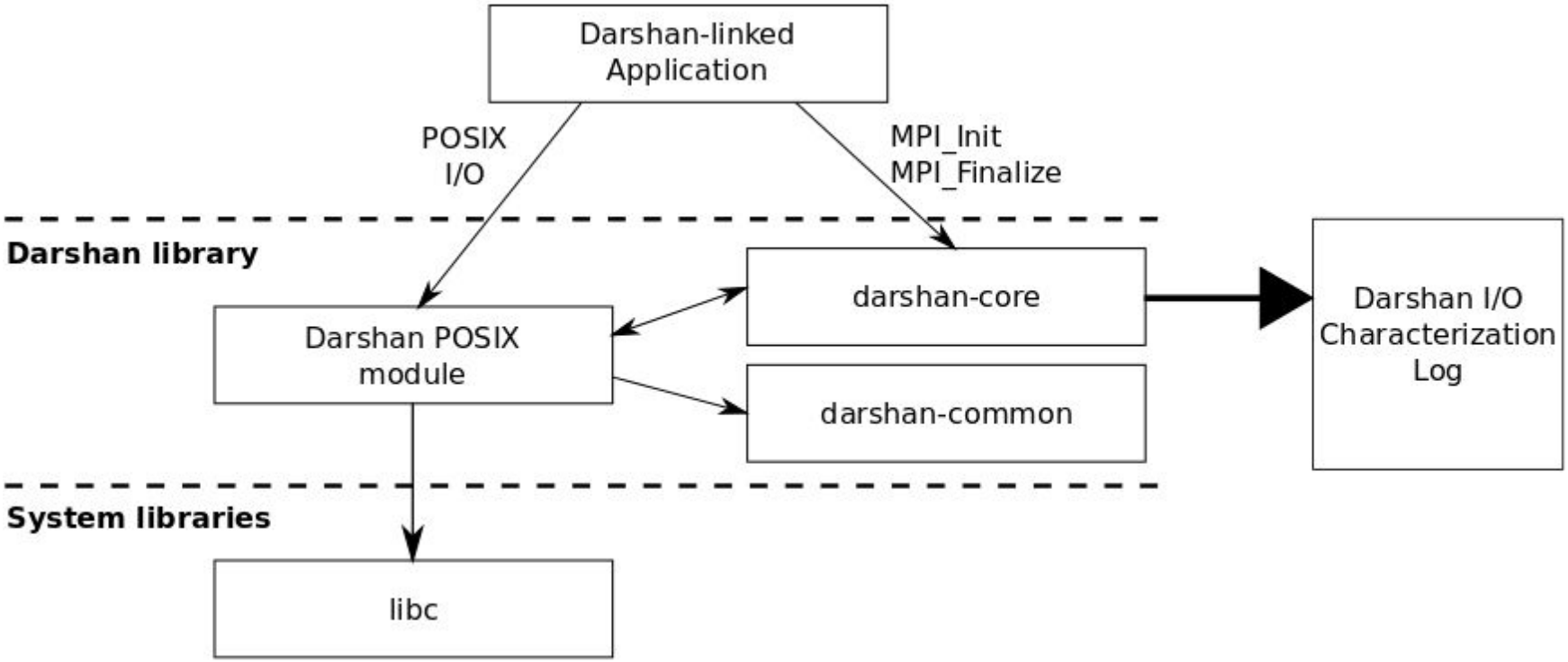
Conjunto de ferramentas para analisar os dados coletados

darshan-runtime

- Coleta informações das aplicações MPI
- Inicia a captura com a chamada MPI_Init()
- Encerra com a chamada MPI_Finalize()
- Captura informações tanto para acessos POSIX quanto MPI-IO
- Informações para HDF5 e PnetCDF são limitadas
- Possui uma API que permite implementar novas funcionalidades

darshan-runtime

LNCC
www.lncc.br
Laboratório Nacional de Computação Científica



darshan-utils

Conjunto de ferramentas para visualizar os resultados coletados:

- darshan-job-summary.pl
- darshan-summary-per-file.sh
- darshan-parser
- darshan-analyser
- darshan-convert

darshan-utils

LNCC
www.lncc.br
Laboratório Nacional de Computação Científica

darshan-job-summary.pl

Gera um resumo, em PDF, sobre as informações do job organizando os dados em tabelas e gráficos.

O arquivo de saída tem o mesmo nome do arquivo do darshan, mas pode ser escolhido um nome diferente com o parâmetro --output

darshan-summary-per-file.sh

Gera um arquivo pdf para cada arquivo que foi acessado durante a execução do job.

darshan-utils

darshan-parser

--all : exibe todas as informações
--base : exibe os dados de log do darshan (default)
--file : exibe informações referentes a todos os arquivos
--file-list : informações de cada arquivo acessado
--perf : informações referentes a performance (unique and shared files)
--total : totalização de todas as estatísticas coletadas

Utilização:

$ darshan-parser [opção] arquivo-de-log.darshan.gz

darshan-utils

LNCC
www.lncc.br
Laboratório Nacional de Computação Científica

darshan-analyzer

Exibe um resumo com o método de acesso usado em todos os arquivos de log de um determinado diretório

log dir: /diretorio/de/arquivos/do/darshan
total logs: 14
        shared file access: 0.853211 [4]
        file-per-proccess access: 0.091743 [2]
        mpio access: 0.009174 [8]
        pnetcdf access: 0.000000 [0]
        hdf5 access: 0.000000 [0]
I/O percentage of runtime:
    0.00-10.00: 9
    10.20-20.00: 5

darshan-utils

LNCC
www.lncc.br
Laboratório Nacional de Computação Científica

darshan-convert

Utilizado para converter arquivos de formatos antigos no formato atual e também permite remover as informações que identificam os arquivos assim como adicionar informações ao cabeçalho dos arquivos de log.

Utilização

Modo instrumentado

Necessita recompilar as aplicações
Utilização de wrappers para os compiladores MPI

Modo não instrumentado

Não é necessário recompilar a aplicação
Funciona com o pré-carregamento da biblioteca através da variável de ambiente LD_PRELOAD

Utilização

Limitação com o uso do Intel MPI

Os módulos do Darshan para Intel MPI (instrumentado e não instrumentado) funcionam apenas com executáveis C em C++, não tendo suporte para executáveis Fortran.

É necessário definir as variáveis de ambiente abaixo:

```
I_MPI_EXTRA_FILESYSTEM=on
I_MPI_EXTRA_FILESYSTEM_LIST=lustre
```

Utilização

Carregar o módulo do darshan

```
module load darshan/3.2.1_openmpi_gnu_4.0.1
```

Definir a variável de ambiente DARSHAN_LOGPATH

Antes de executar a aplicação é necessário configurar a variável de ambiente DARSHAN_LOGPATH com o caminho onde os arquivos de log serão armazenados.

```
export  DARSHAN_LOGPATH=${SCRATCH}/darshan_logs
```

Utilização

Enviar o job para o SLURM.

```
sbatch meuscript.sh
```

Ao final da execução será criado o arquivo com os dados coletados.

Exemplo:

```
${SCRATCH}/darshan_logs/
username_IOR_id108433_8-4-31528-9724920258954479552_1.darshan
```

Gerando um arquivo PDF com as estatísticas de I/O do job

```
$ darshan-job-summary.pl  ${SCRATCH}/darshan_logs/
username_IOR_id108433_8-4-31528-9724920258954479552_1.darshan
```

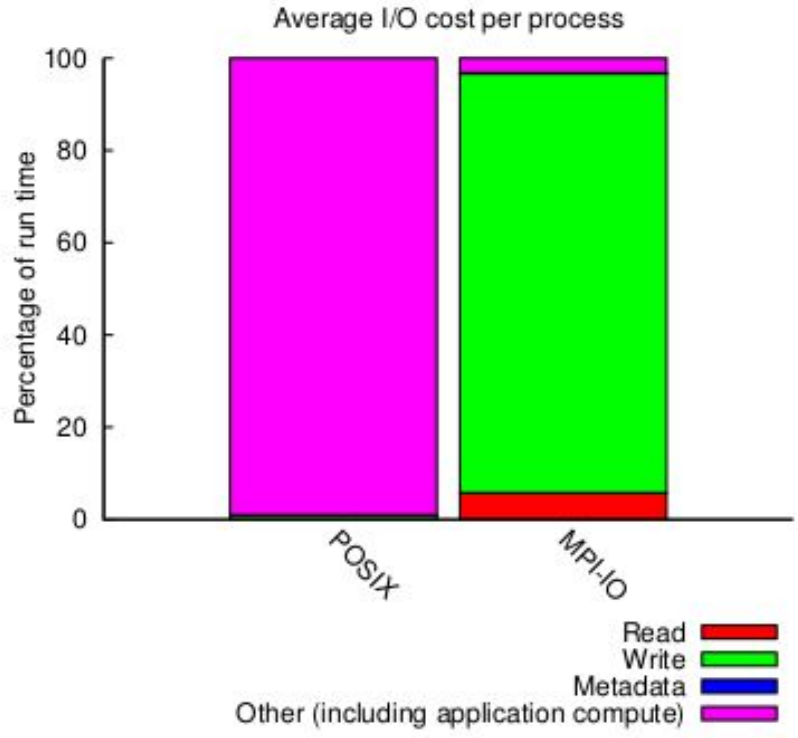
Analisando os resultados

LNCC
www.lncc.br
Laboratório Nacional de Computação Científica

Average I/O cost per process

Quanto tempo da aplicação foi utilizado em operações de I/O em cada API

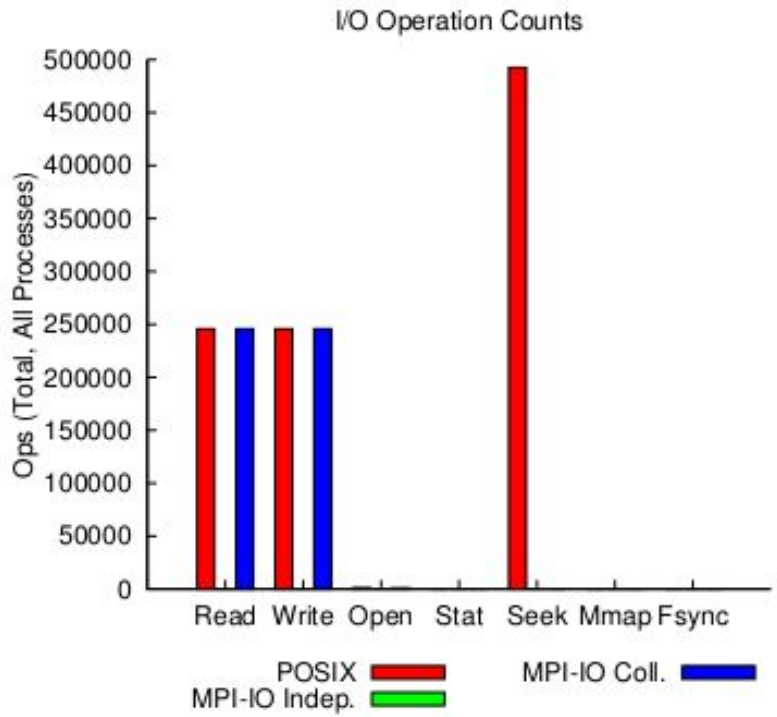
A sobrecarga de acessos a um sistema de arquivos com poucos servidores de metadados pode fazer com que as operações sejam atendidas de forma serial, reduzindo a performance.



Analisando os resultados

I/O Operation Counts

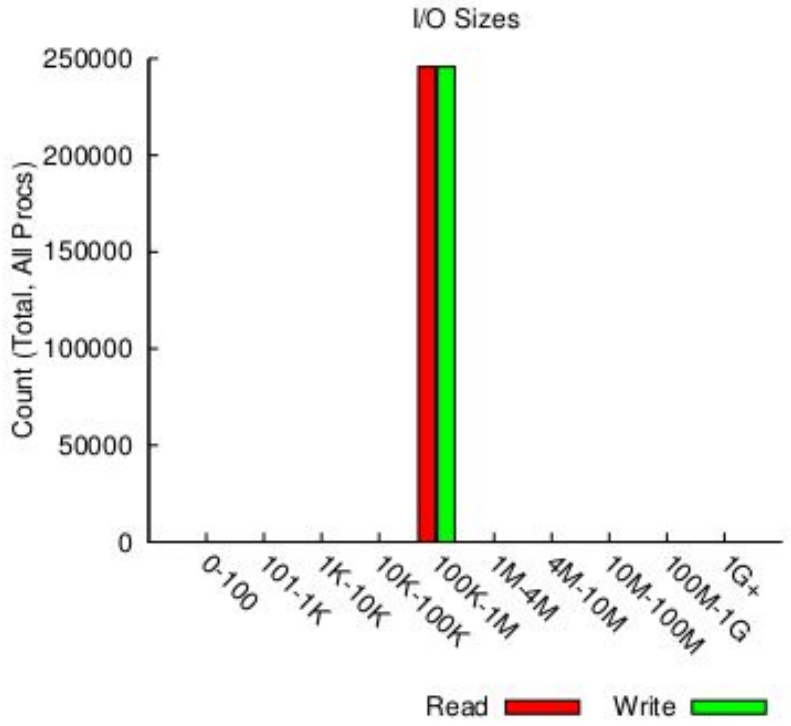
Exibe quantas operações de I/O de cada tipo ocorreram.



Analisando os resultados

I/O Sizes

Exibe o quantidade operações realizadas e seus respectivos tamanhos para leitura e escrita.



Analisando os resultados

I/O Pattern

Informações sobre o padrão de acesso aos arquivos.

Consecutivas:
o acesso aos blocos de dados é feito sem gaps.

Sequencial:
o acesso aos blocos de dados é feito de forma irregular.