

File I/O Patternプロファイル: Darshan

石川裕

理化学研究所 計算科学研究機構

2016/03/23

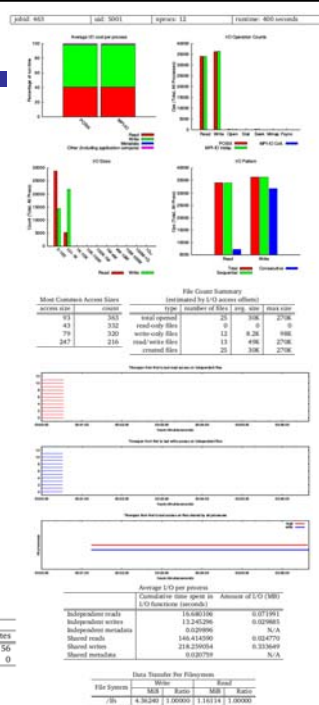
Darshan

- Darshanは米国アルゴンヌ国立研究所で開発されたソフトウェアで以下の特徴を持っています
 - MPIアプリケーションのソースコードを改変することなく各プロセスのファイルI/O処理統計情報を保存
 - ただし、時系列でファイルI/O処理発生時刻を保存してはいない
 - 視覚化ツールが提供されている
- AICSでは以下の改良をしています
 - ファイルI/O処理を時系列で保存
 - Stream I/O系 (fopen/fread/fwrite/fclose)のログも保存
 - MPIアプリケーションでないプログラムのログも保存可能
 - MPIアプリケーション以外に対応するライブラリを提供

- AICS改良版Darshan入手先:

<http://www-sys-aics.riken.jp/releasedsoftware/ksoftware/darshan.html>

File Suffix	Processes	Variance in Shared Files							
		Fastest		Slowest		σ			
		Rank	Time	Rank	Time	Time	Bytes		
.../mpio_resulta	12	1	262.312629	270K	0	262.314721	270K	0.0285	56
.../mpio_resultw	12	2	102.575725	98K	5	102.576477	98K	0.0432	0



2016/03/23

理化学研究所 計算科学研究機構

2

Darshanツールのインストール&使い方

- ホームディレクトリは以下と仮定
 - /home/raXXXXXX/aXXXXX/
- cd ~/work/
- 解凍
 - tar xzf darshan-aics-bin.tar.gz
 - 以下のファイルができる
 - darshan/k-lib/libdarshan.so
 - MPIアプリ用実行時ライブラリ
 - Darshan/k-lib/libdarshan-single.so
 - Non MPIアプリ用実行時ライブラリ
 - darshan/bin/darshan-parser
 - loginノード上でのツール
 - darshan/example/
 - 例
- コンパイル&実行
 - cd ~/work/darshan/example
 - make
 - pjsub ./run-k.sh

/home/raXXXXXX/aXXXXX/work/

```
darshan/
-----
README-aics.txt
k-lib/libdarshan.so
k-lib/libdarshan-single.so
bin/darshan-parser
example/
  Makefile
  run-k.sh
  simple.c
```

README-aics.txtも読んでください

注:AICS改良版darshanは、ファイルシステムマウント情報が正しく取れないため機能を止めています

2016/03/23

理化学研究所 計算科学研究機構

3

実行スクリプト例: MPIアプリケーションの場合

- test/run-k.shを参考に以下のようなジョブスクリプトで実行してください。
- アプリを再コンパイルする必要ありません(static linkしていなければ)。

```
#!/bin/bash
#PJM --rsc-list rscgrp=small
#PJM --rsc-list node=12
#PJM --rsc-list elapse=0:10:00 # HH:MM:SS
#PJM --mpi use-rankdir
#
#PJM --stgin "./simple ."
#PJM --stgin "../darshan/k-lib/libdarshan.so ."
#PJM --stgout-dir "../darshan/ ./results/%n.d%/job/darshan/ stgout=all,recursive=99"
#
./work/system/Env_base

mkdir -p ../darshan/ date +%Y/%m/%d | sed -e "s/¥/0/¥//g"
mpixec -x "LD_PRELOAD=./libdarshan.so" -x "DARSHAN_HISTORY_RW=rw" -ofprefix rank,nid ./simple wdata
```

赤字部分を追加してください

環境変数DARSHAN_HISTORY_RWは、理研計算科学研究機構で拡張した機能を有効にするためのものです

libdashan.soが実行時ワーキングディレクトリ直下にあると想定

YY/MM/DDは実行時の日付、例えば2016年3月23日に実行したのであれば、2016/3/19/

- 実行後、以下のようなディレクトリができる
 - run.sh.d2745580/job/darshan/YY/MM/DD/
- その下に以下のようなファイルができる
 - a03228_simple_id29390_2-24-29839-4689494154855117309_1.darshan.gz

2016/03/23

理化学研究所 計算科学研究機構

4

可視化: darshan-job-summary.pl, darshan-history.pl

- 実行後、以下のようなディレクトリができていると仮定
run.sh.d2745580/job/darshan/2016/3/23/
- その下に以下のようなファイルができていると仮定
a03228_simple_id29390_2-24-29839-4689494154855117309_1.darshan.gz

□ darshan-job-summary.pl

- 以下のコマンドを実行する
\$ cd run.sh.d2745580/job/darshan/2016/3/23/
\$ darshan-job-summary.pl a03228_simple_id29390_2-24-29839-4689494154855117309_1.darshan.gz
- 以下のファイルが生成される。これはdarshanオリジナルの出力
a03228_simple_id29390_2-24-29839-4689494154855117309_1.darshan.gz.pdf

□ darshan-history.pl

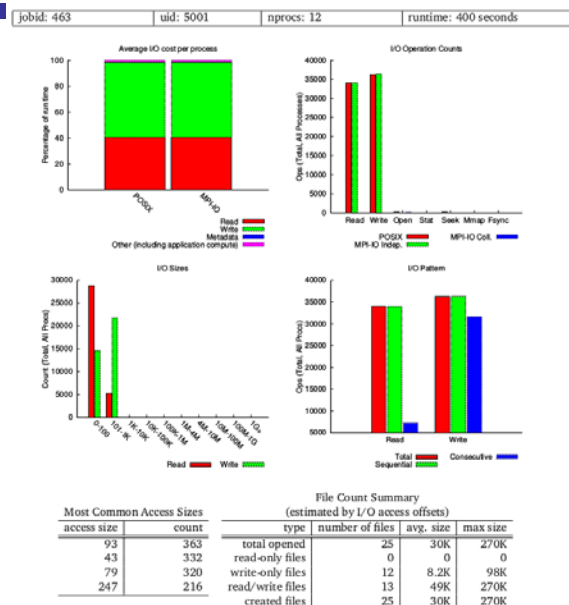
- 実行時に以下の環境変数を設定していた場合
 - DARSHAN_HISTORY_RW=rw
- AICS拡張の時系列ファイルI/Oが記録されているので、以下のコマンドを実行する
\$ darshan-history.pl a03228_simple_id29390_2-24-29839-4689494154855117309_1.darshan.gz
- 以下のファイルが生成される。
history_rank.png

2016/03/23

理化学研究所 計算科学研究機構

5

darshan-job-summary.pl表示例(1/3)

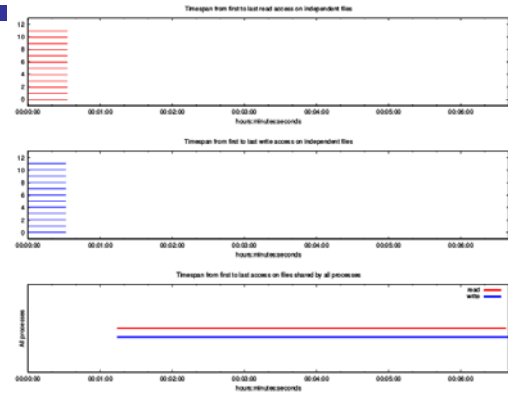


2016/03/23

理化学研究所 計算科学研究機構

6

darshan-job-summary.pl表示例(2/3)



Average I/O per process		
	Cumulative time spent in I/O functions (seconds)	Amount of I/O (MB)
Independent reads	16.680106	0.071991
Independent writes	13.245296	0.029885
Independent metadata	0.029896	N/A
Shared reads	146.414590	0.024770
Shared writes	218.259054	0.333649
Shared metadata	0.020759	N/A

File System	Write		Read	
	MB	Ratio	MB	Ratio
/ifs	4.36240	1.00000	1.16114	1.00000

2016/03/23

理化学研究所 計算科学研究機構

7

darshan-job-summary.pl表示例(3/3)

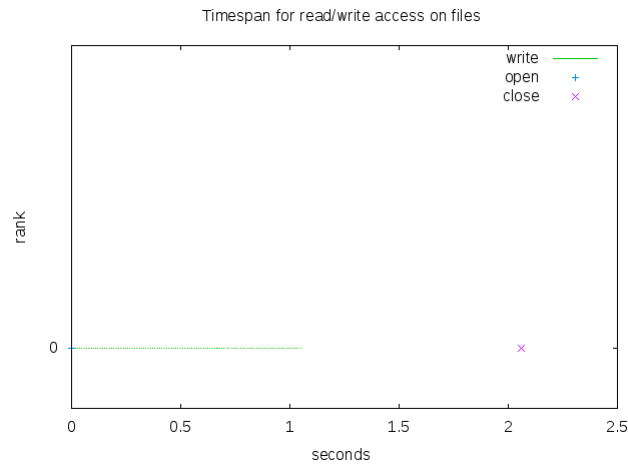
Variance in Shared Files									
File Suffix	Processes	Fastest			Slowest			σ	
		Rank	Time	Bytes	Rank	Time	Bytes	Time	Bytes
...s/mpiio_resulta	12	1	262.312629	270K	0	262.314721	270K	0.0285	56
...s/mpiio_resultw	12	2	102.575725	98K	5	102.576477	98K	0.0433	0

2016/03/23

理化学研究所 計算科学研究機構

8

darshan-history.pl表示例



2016/03/23

理化学研究所 計算科学研究機構

9

実行スクリプト例: Non MPIアプリケーションの場合

```
$ ( export LD_PRELOAD=./libdarshan-single.so; ¥
  export DARSHAN_HISTORY_RW="rw"; ¥
  a.out )
```

- 実行後、以下のようなファイルができる
darshanlog-ubuntu-1458387555-a.out-12760.gz
ログフォーマットは、
 - darshanlog-<host name>-<start time>-<command name>-<pid>.gz
- 可視化のやり方はMPIアプリケーションと同じ

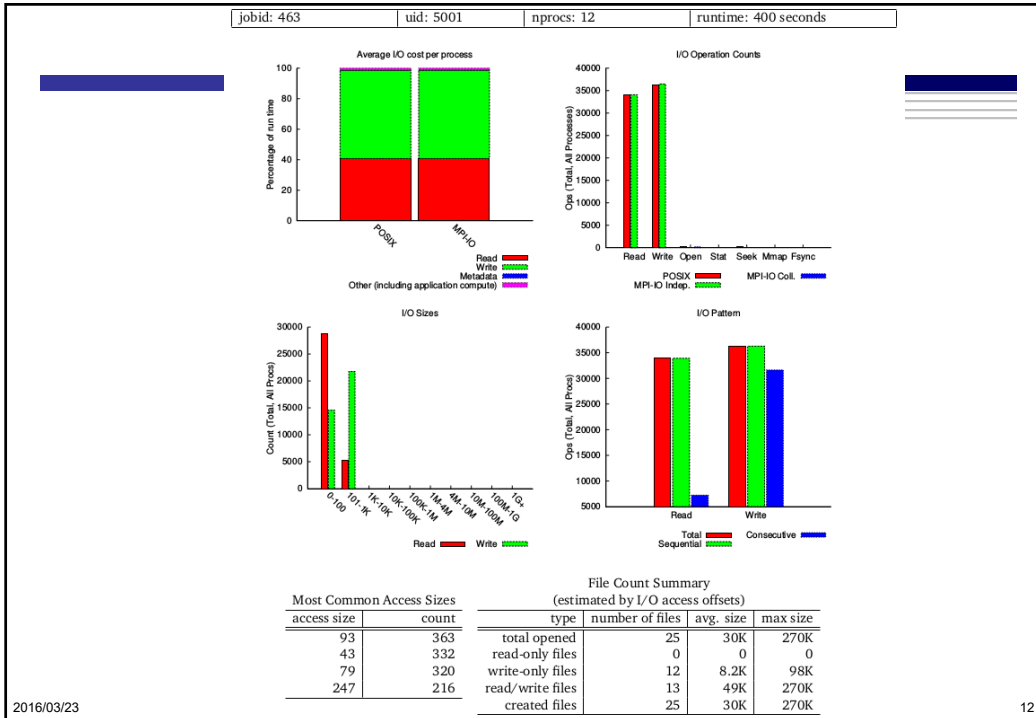
2016/03/23

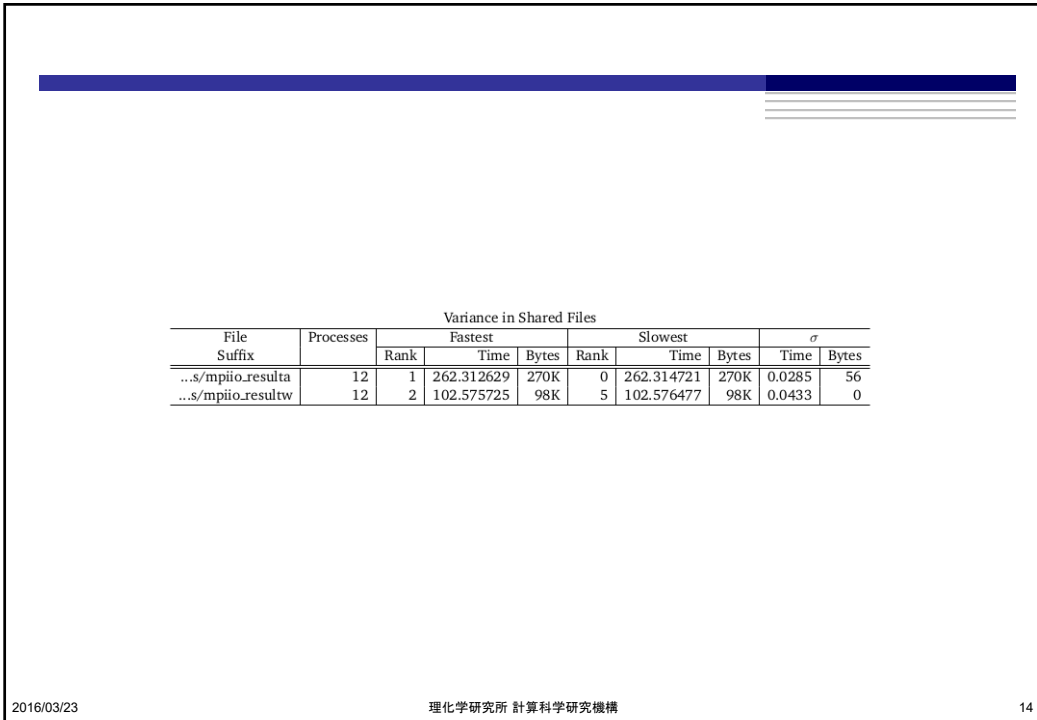
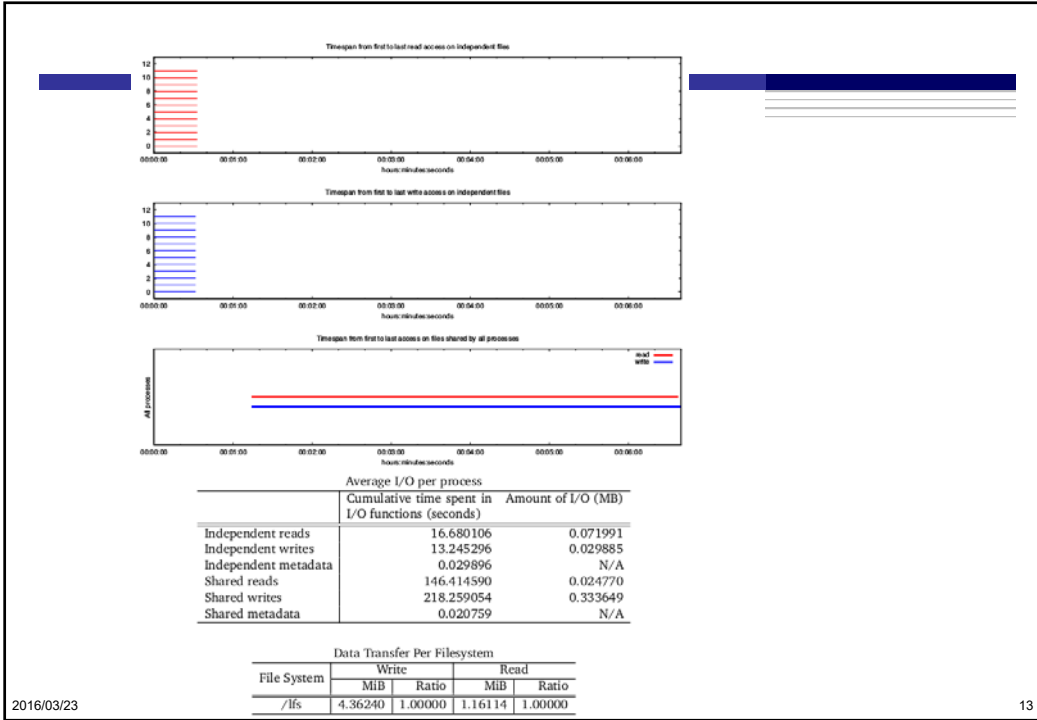
理化学研究所 計算科学研究機構

10

可視化

- 実行後、以下のようなディレクトリができていると仮定
run.sh.d2745580/job/darshan/2016/3/23/
- その下に以下のようなファイルができていると仮定
a03228_simple_id29390_2-24-29839-4689494154855117309_1.darshan.gz
- 以下のコマンドを実行する
\$ cd run.sh.d2745580/job/darshan/2016/3/23/
\$ darshan-job-summary.pl a03228_simple_id29390_2-24-29839-4689494154855117309_1.darshan.gz
- 以下のファイルが生成される
a03228_simple_id29390_2-24-29839-4689494154855117309_1.darshan.gz.pdf
- pdf viewerで上記ファイルを見る





darshan-parser実行例

```
$ darshan-parser a03228_simple_id29390_2-24-29839-4689494154855117309_1.darshan.gz
```

```
# darshan log version: 2.05h
# size of file statistics: 1328 bytes
# size of job statistics: 1080 bytes
# exe: ./simple wdata
# uid: 38127
# jobid: 29302
# start_time: 1429094485
# start_time_ascii: Wed Apr 15 19:41:25 2015
# end_time: 1429094491
# end_time_ascii: Wed Apr 15 19:41:31 2015
# nprocs: 12
# run time: 7
# metadata: lib_ver = 2.3.0
# metadata: h = romio_no_indep_rw=true;cb_nodes=4
# <0> write mple/wdata0-a-0
# start = 1404576.179318 count=90
# time, elapsed, Kbyte
0.000000, 0.001545, 0.100000
0.002586, 0.000046, 0.100000
0.002638, 0.000042, 0.100000
.....
```

2016/03/23

理化学研究所 計算科学研究機構

15

darshanソースをコンパイルしたい人向け情報

- ✓ ソースプログラムは以下から入手可能

<http://www.mcs.anl.gov/research/projects/darshan/>

- ✓ Configure例

Kの場合:

```
./configure --host=x86_64-unknown-linux-gnu ¥
--target=sparc64v-sparc-linux-gnu --with-mem-align=8 ¥
--with-log-path=./darshan --with-jobid-env=NONE ¥
CC="mpicc -Xg" ¥ このオプション重要(gnuコンパチ)
CFLAGS="-O2 -g -DHISTORY"
```

Linux-Xeon例:

```
./configure --with-mem-align=8 ¥
--with-log-path=./darshan ¥
--with-jobid-env=PBS_JOBID CC=mpicc ¥
CFLAGS="-O2 -g -DHISTORY"
```

2016/03/23

理化学研究所 計算科学研究機構

16